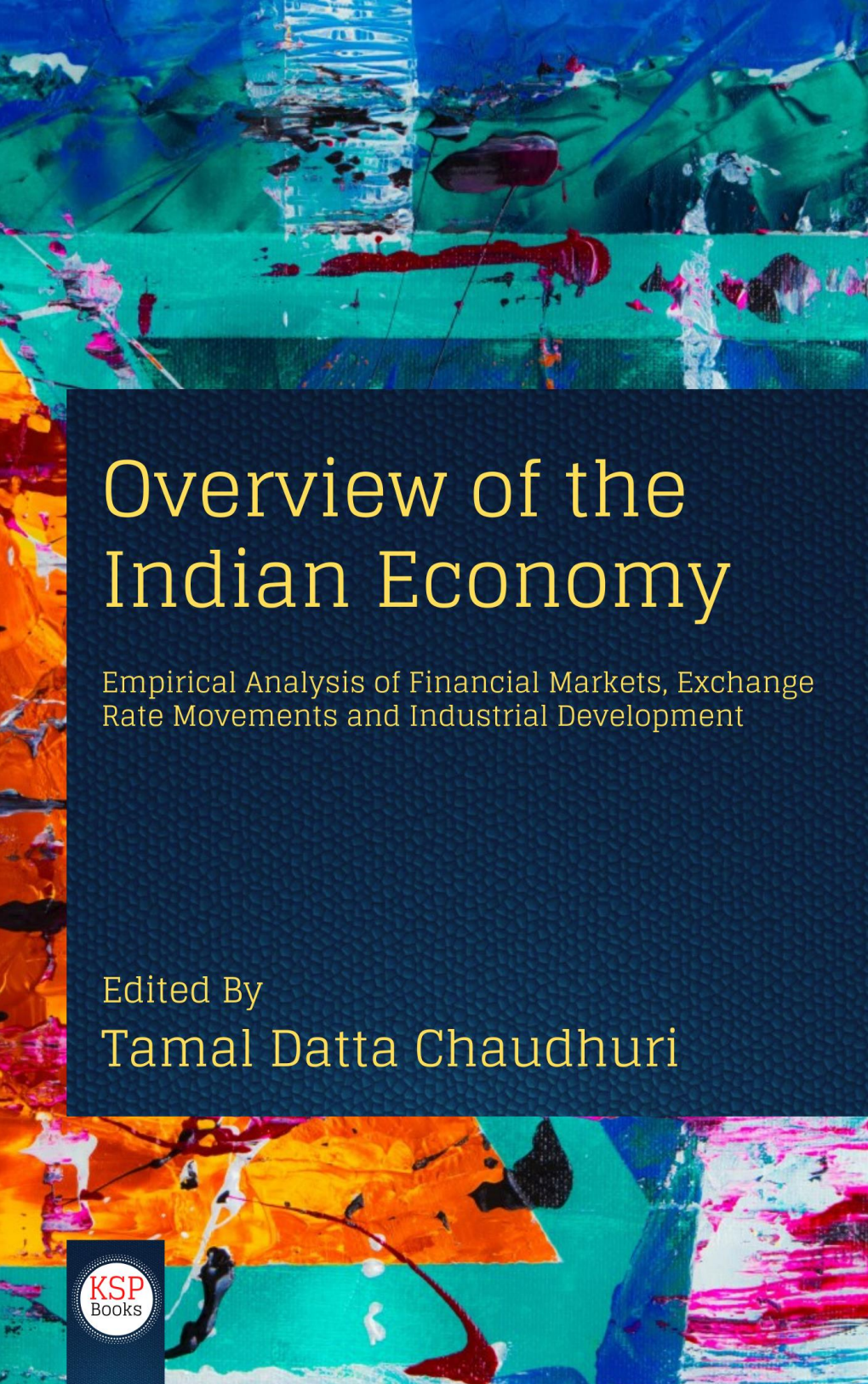

# Overview of the Indian Economy

## Empirical Analysis of Financial Markets, Exchange Rate Movements and Industrial Development

Edited By

Tamal Datta Chaudhuri

KSP Books

# Overview of the Indian Economy

## Empirical analysis of financial markets, exchange rate movements and industrial development

*Edited by*
**Tamal Datta Chaudhuri**
Calcutta Business School, India

KSP Books
econsciences.com

# Overview of the Indian Economy

## Empirical analysis of financial markets, exchange rate movements and industrial development

*Edited By*
Tamal Datta Chaudhuri

KSP Books
econsciences.com

ISBN: 978-605-7736-93-2 (e-Book)
KSP Books 2020

***Overview of the Indian Economy:*** *Empirical analysis of financial markets, exchange rate movements and industrial development*

Editor: ***Tamal Datta Chaudhuri***

Calcutta Business School, Diamond Harbor Road, Bishnupur -743503, 24 South Parganas, West Bengal, India.

© KSP Books 2020

**Open Access** This book is distributed under the terms of the Creative Commons Attribution-Noncommercial 4.0 IGO (CC BY-NC 4.0 IGO) License which permits any noncommercial use, distribution, and reproduction in any medium, provided ADB and the original author(s) and source are credited.

Open Access This book is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited. All commercial rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for commercial use must always be obtained from KSP Books. Permissions for commercial use may be obtained through Rights Link at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law. The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use. While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

This book licensed under Creative Commons Attribution-NonCommercial license (4.0)

econsciences.com

# Preface

This book brings together papers which highlight various aspects of development of the Indian economy in the recent past. It covers Indian financial markets, movements in the exchange rate, and value creation and innovativeness of the manufacturing sector. In the various chapters, application of analytical tools like R, Artificial Neural Network (ANN), Clustering are demonstrated. While two papers dwell on sectoral characteristics and portfolio choice for value creation, two papers focus on understanding stock market volatility and the sources of volatility.

For understanding industrial development, economic profit generated by Indian manufacturing companies are calculated over two time periods to see whether there has been any fundamental change in their performance and the sectors within which they belong. The purpose is two-fold. First, to get an idea about how Indian companies have fared

over the two time periods and whether there has been any structural change. Second, help companies decide on their next strategic move and allocate funds for the purpose.

The reasons for studying stock market volatility are that it i) aids in intraday trading, ii) is the basis of neutral trading in the options market, iii) affects portfolio rebalancing by fund managers, iv) helps in hedging, v) affects capital budgeting decisions through timing of raising equity from the market and its pricing and also vi) affects policy decisions relating to the financial markets. In todays globalized environment, with increased financial integration and also enhanced trade in goods and services, volatility in one country spreads to other countries almost immediately. In India, where foreign institutional investors (FIIs) are large players in the stock market, their fund allocation is shaped by macroeconomic conditions in other economies. Thus any macroeconomic event in any part of the world causes reallocation of FII funds, leading to volatility in Indian stock markets.

Any discussion on exchange rate movements and forecasting should include explanatory variables from both the current account and the capital account of the balance of payments. In our study we include such factors to forecast the value of the Indian rupee vis a vis the US Dollar. Further, factors reflecting political instability and lack of mechanism for enforcement of contracts that can affect both direct foreign investment and also portfolio investment have been incorporated.

**T.D. Chaudhuri**
March 14, 2020

# Notes on Contributors

Tamal Datta **Chaudhuri.** Professor Tamal Datta Chaudhuri did his graduation in Economics from Presidency College, Calcutta and his MSc in Economics from Calcutta University. He obtained his PhD in Economics from The Johns Hopkins University, USA. He taught in Calcutta University for 15 years and then joined a Development Financial Institution, Industrial Investment Bank of India, where he became Chief General Manager in Charge. He has 42 years of teaching experience and 23 years of industry experience. He has taught in several institutes like Indian Statistical Institute, Calcutta, ICFAI Business School, Calcutta and Hyderabad, IISWBM, IIFT, IMT Ghaziabad and Towson State University, Maryland, USA. He was Visiting Fellow at the University of Illinois, Urbana Champaign, USA. The subjects that he has taught are Microeconomics, Macroeconomics, Critical Analysis of Organizations, Stock Market Simulation Game, Concepts in Management, Security Analysis and Portfolio Management, Corporate Finance, Financial Derivatives Management, Risk Management in Banks and Treasury and Foreign Exchange Management. He has published several research papers in

domestic and foreign journals and has a few books to his credit. Dr. DattaChaudhuri has executed research projects on behalf of the Planning Commission and the Department of Science and Technology, Ministry of Science and Technology, Government of India. He has been an invitee to the Salzburg Seminar, Austria. He is currently Professor and Dean of Calcutta Business School, Calcutta, India.

Indrail **Ghosh.** Professor Indranil Ghosh completed his graduation and post-graduation in Information Technology and Industrial Engineering & Management from West Bengal University of Technology (Currently known as MaulanaAbulKalam Azad University of Technology). He is currently working as an Assistant Professor in the area of Decision Science and Systems Management at Calcutta Business School, Calcutta. His research areas includes Machine Learning, Pattern Recognition, Modelling Financial Markets, etc. Previously he had worked as Systems Engineer at Infosys Limited, India and as project linked person at Indian Statistical Institute. He has published many papers in national and international journals and has a few books to his credit.

Jaydip **Sen.** Professor Jaydip Sen is Professor and Head of the School of Computing, Analytics Cyber Security and Applied Sciences in NSHM College of Management and Technology in Kolkata. Prior to this, he was Professor of Information Technology and Analytics Department in Praxis Business School, Kolkata. His primary research areas include Security and Privacy Issues in Computing and Communications, Internet of Things, Wireless Sensor Networks, Machine Learning and Artificial Intelligence and Time Series Analysis and Forecasting of Financial Data and Deep Learning. He has published many papers in national and international journals and has many books to his credit. His research work has over 3800 citations.

Aditya **Jhunjhunwala**. Aditya Jhunjhunwala holds a Post Graduate Diploma in Management from Calcutta Business School. He is currently employed as Associate, PricewaterhouseCoopers: PwC SDC, Kolkata.

Gulshan K. **Bhamrah.** Gulshan Kaur Bhamrah holds a Post Graduate Diploma in Management from Calcutta Business School. She is currently employed as Audit Associate, KPMG, Bangalore.

# Acknowledgements

*Students have always been my inspiration for research and many of the ideas presented in this book took shape in various classes on finance that I taught in Calcutta Business School and IBS Kolkata. Along the way, some students have collaborated with me in my research, and this volume has two papers written with two students of mine, Aditya Jhunjhunwala and Gulshan Kaur Bhamrah. I was quite excited about the way they went about their research and gave shape to the thoughts. I thank them for their interest and collaboration.*

*Professor Jaydip Sen is an experienced researcher and we found some common areas where we could collaborate when he was teaching in Calcutta Business School. His deep knowledge of machine learning and deep learning tools helped provide thorough analysis of stock price movements and sectoral characteristics of the Indian stock market. I thank him for contribution.*

*I work very closely with Professor Indranil Ghosh who has been associated with me in Calcutta Business School for quite some time. Among the different courses he teaches, his course on Financial Analytics is highly appreciated. He is a serious researcher and has presented papers in reputed conferences in India. This volume presents some papers that we have written together and I thank him for his efforts in giving shape to the papers*

# Contents

Introduction 1

Chapter I

Value creation by Indian companies: A comparative study over two time periods

Aditya Jhunjhunwala, Tamal D. Chaudhuri & Gulshan K. Bhamrah

*7-30*

Objective 8
Literature review 9
Methodology 11
Data and resultes 12
Concluding remarks 27
References 29

## Chapter 2

### Can portfolio returns exceed market return? An examination of the efficient market hypothesis for the Indian stock market

Tamal D. Chaudhuri & Gulshan K. Bhamrah

*31-44*

Introduction 30
Literature review 33
Methodology 35
Results 36
Conclusion 41
References 43

## Chapter 3

### An alternative framework for time series decomposition and forecasting and its relevance for portfolio choice - A comparative study of the Indian consumer durable and small cap sectors

Jaydip Sen & Tamal D. Chaudhuri

*45-87*

Introduction 45

Methodology 48
Time series decomposition results 49
Proposed frameworks of time series forecasting 59
Forecasting results and analysis 63
Related work 79
Conclusion 82
References 84

## Chapter 4

### Using clustering method to understand Indian stock market volatility

Tamal D. Chaudhuri & Indranil Ghosh

*88-112*

Introduction 88
Objective of the chapter 90
Methodology 91
**Kernel K-Means 92**
**Gaussian mixture model 93**
**Self-organizing map 96**
**Silhouette index (SI) 98**
**Dunn index (DI) 98**
Literature review 98
The variables 100
Results and analysis 102
Concluding remarks 108
References 110

## Chapter 5

### Forecasting volatility in Indian stock market using artificial neural network with multiple inputs and outputs

Tamal D. Chaudhuri & Indranil Ghosh

*113-140*

Introduction 113
Objective of the chapter 115
Methodology 116
Literature review 119
The variables 122
Results and analysis 126
Concluding remarks 136
References 137

## Chapter 6

## Artificial neural network and time series modeling based approach to forecasting the exchange rate in a multivariate framework

Tamal D. Chaudhuri & Indranil Ghosh

*141-178*

Introduction 141
Literature review 145
Data and methodology 149
**Multilayer Feed-Forward Network (MLFFNN) 150**
**NARX (Nonlinear Autoregressive models with exogenous input) Neural Network Model 153**
**Time Series Modelling 155**
**GARCH 156**
**EGARCH 157**
**Unit Root Test 157**
Results and based modeling 158
**Results of MLFFNN modelling 158**
**Results of NARX modelling 161**
**Assesment of parametres 164**
Results of time series based modeling 165
Comparative analysis 172
Concluding remarks 173
References 175

Chapter 7

# A predictive analysis of the Indian FMCG sector using time series decomposition - based approach

Jaydip Sen & Tamal D. Chaudhuri

*179-219*


| | |
|---|---|
| Introduction | 179 |
| Methodology | 181 |
| Time series decomposition results | 183 |
| Proposed forecasting methods | 190 |
| Forecasting results | 194 |
| Summary of forecasting results | 206 |
| Related work | 208 |
| Conclusion | 212 |
| References | 214 |

# Introduction

This book is a collection of essays on the Indian economy. It focuses on industrial development, stock markets, the exchange rate, volatility movements and sectoral characteristics of the Indian economy. The objective of Chapter 1 titled "Value creation by Indian companies: A comparative study over two time periods" is to derive economic profit generated by Indian companies over two time periods and see whether there has been any fundamental change in the performance of companies and the sectors within which they belong. The focus is on non-finance companies. The purpose is two-fold. First, to get an idea about how Indian companies have fared over the two time periods and whether there has been any structural change. Second, to help companies decide on their next strategic move and allocate funds for the purpose. The study also focusses on the relationship between size and economic profit, where invested capital and market capitalization

represents size. The methodology presented in the chapter enables us to understand the performance of Indian companies and also the sectors within which they belong.

Chapter 2 titled "Can portfolio returns exceed market return? An examination of the efficient market hypothesis for the Indian stock market" explores the possibility of forming portfolio of stocks that can generate returns higher than the market over a time period. Various principles are used for portfolio formation in the year 2013, and it is examined whether such portfolios have been able to generate excess returns over the next five years. Data has been used for Indian companies which are listed in the National Stock Exchange and Bombay Stock Exchange. The period under consideration has seen upswings and downswings, and the paper explores whether the portfolios have been able to generate excess returns.

One of the challenging research problems in the domain of time series analysis and forecasting is making efficient and robust prediction of stock market prices. With rapid development and evolution of sophisticated algorithms and with the availability of extremely fast computing platforms, it has now become possible to effectively extract, store, process and analyze high volume stock market time series data. Complex algorithms for forecasting are now available for speedy execution over parallel architecture leading to fairly accurate results. Chapter 3 titled "An alternative framework for time series decomposition and forecasting and its relevance for portfolio choice - A comparative study of the Indian consumer durable and small cap sectors" proposes a decomposition approach for better understanding of the behavior of each of the time series. The contention is that various sectors reveal different time series patterns and understanding them is essential for portfolio formation. Based on this structural analysis, some robust forecasting techniques are used and their efficiency are

analyzed by their accuracy in prediction using suitably chosen training and test data sets.

The basic proposition of Chapter 4 titled "Using clustering method to understand Indian stock market volatility" is whether stock market volatility can be predicted at all, and if so, when it can be predicted. The exercise has been performed for the Indian stock market on daily data for two years. For the analysis, number of clusters are mapped against number of variables. The contention is that, given a fixed number of variables, one of them being historic volatility of NIFTY returns, if increase in the number of clusters improves clustering efficiency, then volatility cannot be predicted. Volatility then becomes random as, for a given time period, it gets classified in various clusters. On the other hand, if efficiency falls with increase in the number of clusters, then volatility can be predicted as there is some homogeneity in the data. If the number of clusters are fixed and then the number of variables are increased, this should have some impact on clustering efficiency. Indeed if we can hit upon, in a sense, an optimum number of variables, then if the number of clusters is reasonably small, we can use these variables to predict volatility. The variables that are considered for the study are volatility of NIFTY returns, volatility of gold returns, India VIX, CBOE VIX, volatility of crude oil returns, volatility of DJIA returns, volatility of DAX returns, volatility of Hang Seng returns and volatility of Nikkei returns. Three clustering algorithms are used namely Kernel K-Means, Self-Organizing Maps and Mixture of Gaussian models and two internal clustering validity measures are applied, namely Silhouette Index and Dunn Index, to assess the quality of generated clusters.

Volatility in stock markets has been extensively studiedin the applied finance literature. In Chapter 5 titled "Forecasting volatility in Indian stock market using artificial neural network with multiple inputs and outputs", Artificial

Neural Network models based on various back propagation algorithms have been constructed to predict volatility in the Indian stock market through volatility of NIFTY returns and volatility of gold returns. This model considers India VIX, CBOE VIX, volatility of crude oil returns (CRUDESDR), volatility of DJIA returns (DJIASDR), volatility of DAX returns (DAXSDR), volatility of Hang Seng returns (HANGSDR) and volatility of Nikkei returns (NIKKEISDR) as predictor variables. Three sets of experiments have been performed over three time periods to judge the effectiveness of the approach.

Any discussion on exchange rate movements and forecasting should include explanatory variables from both the current account and the capital account of the balance of payments. In Chapter 6 titled "Artificial neural network and time series modeling based approach to forecasting the exchange rate in a multivariate framework"such factors are included to forecast the value of the Indian rupee vis a vis the US Dollar. Further, factors reflecting political instability and lack of mechanism for enforcement of contracts that can affect both direct foreign investment and also portfolio investment, have been incorporated. The explanatory variables chosen are the 3 month Rupee Dollar futures exchange rate (FX4), NIFTY returns (NIFTYR), Dow Jones Industrial Average returns (DJIAR), Hang Seng returns (HSR), DAX returns (DR), crude oil price (COP), CBOE VIX (CV) and India VIX (IV). To forecast the exchange rate, two different classes of frameworks have been used namely, Artificial Neural Network (ANN) based models and Time Series Econometric models. Multilayer Feed Forward Neural Network (MLFFNN) and Nonlinear Autoregressive models with Exogenous Input (NARX) Neural Network are the approaches that have been used as ANN models. Generalized Autoregressive Conditional Heteroskedastic (GARCH) and Exponential Generalized Autoregressive

Conditional Heteroskedastic (EGARCH) techniques are the ones that have been used as Time Series Econometric methods.

**T.D. Chaudhuri**
March 14, 2020

# 1 Value creation by Indian companies: A comparative study over two time periods

*By*
Aditya JHUNJHUNWALA
Tamal D. CHAUDHURI
& Gulshan K. BHAMRAH

## Introduction

Companies evolve over time. Some grow and become industry leaders, some remain niche players, some go transnational, while some find it difficult to compete and stagnate. While there are industry level factors that shape the fortune of a company, it is also internal factors that significantly matter. There are books written on how companies become successful, and to name a few we have In Search of Excellence by Peters & Waterman (1982), Built to Last by Collins & Porras (1994), Blue Ocean Strategy by Kim & Mauborgne (2005), The End of Competitive Advantage by McGrath (2013) and 3 Box Strategy by Govindrajan (2016). These books draw their views from observing companies over time and look at their historic background and growth process. A recent book titled Strategy Beyond the Hockey Stick by Bradley, Hirt & Smit (2018) provide a yet interesting approach to understanding performance of companies. By evaluating companies on the basis of economic profit, they

identify that whatever be the strategic decisions taken by a company for maintaining competitive advantage, it is the dynamics of different sectors that play a crucial role in shaping the future of a company. They advance the hypothesis that in the overall scheme of things, companies tend to be myopic in their approach and put too much emphasis on self-belief while designing business plans and strategic plans. The book demonstrates that many of the plans may not be successful, not because they are ill-conceived, but because the sector is overall not positioned well.

## Objective

The overall economic performance of an economy and its growth prospects depend on the performance of various sectors and companies belonging to the sectors, the domestic market conditions, the global economic environment and appropriateness of domestic policy measures. Every political regime undertakes numerous policy measures and they try to highlight the success of these policy measures. Many of these measures are forward looking, while some are in response to specific events. Many a times policies do not yield the desired results as they were not well conceived, or were not properly executed, or the external environment was not conducive. Whatever be the reasons, economic performance of an economy affects employment prospects and growth in physical and financial assets. The objective of this chapter is to look at economic profit generated by Indian companies over two time periods and see whether there has been any fundamental change in the performance of companies and sectors within which they belong. We focus on non-finance companies. The purpose is two-fold. First, to get an idea about how Indian companies have fared over the time period and whether there has been any structural

change. Second, to help companies decide on their next strategic move and allocate funds for the purpose.

The plan of the chapter is as follows. A brief literature survey is presented in Section 3. The methodology followed in the study is laid out in Section 4. Section 5 presents the data and the results. Section 6 concludes the chapter.

## Literature review

Porter (1985) noted that "…Not all industries offer equal opportunities for sustained profitability, and the inherent profitability of its industry is one essential ingredient in determining the profitability of a firm. …All industries are not alike from the standpoint of inherent profitability. In industries where the five forces are favorable such a pharmaceuticals, soft drinks, and data base publishing, many competitors earn attractive returns; but in industries where pressure from one or more of the forces is intense, such as rubber, steel, and video games, few firms command attractive returns despite the best efforts of management." There is explicit recognition in the above statements that in order to merely survive, in some industries, firms may have to apply themselves to the fullest extent and take cognizance of the five forces in their strategic thinking. In some industries, profit generation may be relatively easier, given the opportunities that the industry provides.

One way of understanding a company and an industry is through a process of extensive research. Peters & Waterman (1982) is based on structured interviews and also study of annual reports and press clippings of seventy five highly regarded companies. Their book was focused on finding common traits among successful companies and chose them from a wide variety of industries for proper representation.

In a similar vein, Collins & Porras (1994) analyze reasons behind success and failure of companies from similar industries. Thus, industry per se is not the focus of the book.

Rather it is the company. However, industries move forward through the performance of its constituent firms, who in turn compete with each other in terms of value offering, price, distinctiveness, leadership quality and strategic orientation.

This can be observed in Govindrajan (2016) where he advances a simple framework involving Forget the Past, Manage the Present, and Create the Future. Although he focusses on individual companies, industry level factors like allowing new ideas and new trends, sensitivity to regulatory changes, effects of disruptive technologies and new distribution channels are also addressed. The success of a company depends on both external industry specific and industry-wide factors and internal factors. It is ability to adapt and manage that leads to success.

McGrath (2013) develops the concept of Transient Advantage. Her thesis is that companies which try to survive by exploiting competitive advantage may not survive for long as the industry scenario has become dynamic and competitive advantage is transient in nature. Companies need to be nimble, forward thinking, innovative and open to disengage. She advises companies to look out for signs of diminishing returns to innovation, increasing commoditization and diminishing returns to capital. One of the important elements of her suggestion to companies is to aggressively focus on developments in the external world.

This last thought has been given shape and comprehensively dealt with in Bradley *et al.*, (2018). According to their study, strategy by companies generally boils down to repeating whatever the company was doing in the past. This is a result of behavioral and social factors like halo bias, anchoring, confirmation bias, champion bias and loss aversion. Many a times companies start out with great plans that require big funding, only to see the funding thinly spread across existing activities as the management was unable to gamble with the unknown. The authors advance

the concept of a hockey stick to describe a strategic plan, and point out that such strategies were all inward looking, rather than being anchored in external developments in the market place. They then construct a Power Curve to point out that not every industry is positioned to generate significant economic profit, and hence thinking in terms of a hockey stick that after initial losses, every strategy will fly, will not work. Analyzing the external environment and studying other industries is essential for strategy formulation.

Kim & Mauborgne (2005) advance the hypothesis that competing with rivals is not the way to survive and grow. This according to them is like swimming in a red ocean. The strategy should be to make the competition irrelevant by identifying areas of operation where no one has gone before. For this they advance a Four Actions Framework and suggest that companies look across alternative industries. They have a separate chapter advising to look at the bigger picture and not the numbers.

Such a suggestion is present in Porter (1996). He draws a possibility frontier between non-price value delivered and relative cost position. The essence of the frontier is that if there is low non-price value of an offering, then to survive in the market a company has to be cost effective. Strategies aimed at cost reduction and operational efficiency is required in red oceans. If non-price value is delivered, then cost efficiency and pricing is not that important. Further, for a product or a service to be unique, the strategy should be so ring fenced that the processes cannot be imitated in totality.

## Methodology

This study is based on the performance of 3060 Indian non-finance companies over two time periods 2011-2013 and 2014-2017. The data has been sourced from CMIE Prowess Data Base. We look at the overall performance of these companies, performance of the sector/industry where they

belong, and also their performance in terms of size measured by market capitalization. We further focus on 45 large cap companies, 31 mid cap companies and 496 small companies. These are non-finance companies from the BSE large cap, mid cap and small cap indices. Our study also looks at performance of 20 sectors to which the above 3060 companies belong.

The metric that we consider for measuring performance is economic profit. Economic profit is arrived at after subtracting the opportunity cost of capital from operating profit. Operating profit divided by capital employed gives the return from capital employed. By subtracting the opportunity cost of capital from this, we arrive at the rate of economic profit. This, multiplied by capital employed, gives the level of economic profit. This value we compute for companies from different sectors and various levels of market capitalization to arrive at our results.

For proper comparison, we have taken data on companies which were in operation in both time periods.

As per Reserve Bank of India website [Retrieved from], Bank Group-wise Weighted Average Lending Rate during 2011 to 2013 was around 10.50%. This fell to around 9.20% in October 2018. According to State Bank of India information, their Benchmark Prime Lending Rate was around 11% to 14% during 2011-13, and was 13.40% during 2017. Given these rates, and given the fact that borrowing rates for companies depends on also their credit rating, we have considered 12% to be the opportunity cost of capital for both periods.

## Data and results

Figure 1 shows economic profit for 3060 non finance companies for the period 2011 – 2013. It is drawn by taking the average over 3 years and over 3060 companies for invested capital and operating profit. Table 1 indicates that

the average returns on invested capital for the period was 13.98%. Given opportunity cost at 12%, rate of average economic profit was 1.98% during the period.

For the period 2014 – 2017, as shown in Table 2, the average returns on invested capital turns out to be 12.40%. Given opportunity cost of capital at 12%, rate of average economic profit was .40%. Figure 2 presents the data for the time period. The figures show, that although average rate of economic profit has declined, more companies during 2014-17 have broken away from the pack and shown improved performance.

Comparing the two time periods, we observe that there has been an overall decline in the average performance of the sample set of companies, although there are some outliers. It would be interesting to note a) which are the companies that have performed well; b) which are the companies that have performed poorly; c) whether there has been any change in the performance of specific companies; d) which are the industries to which these companies belong; and e) whether there has been a change in the relative position of the sectors.

**Table 1.** *3060 companies for 2011-2013*

| | |
|---|---|
| Average Operating Profit (Rs. Crore) | 150.86 |
| Average Invested Capital (Rs. Crore) | 1078.80 |
| Opportunity Cost of Capital | 12% |
| Average Return on Invested Capital | 13.98% |

**Table 2.** *3060 companies for 2014-2017*

| | |
|---|---|
| Average Operating Profit (Rs. Crore) | 188.57 |
| Average Invested Capital (Rs. Crore) | 1520.95 |
| Opportunity Cost of Capital | 12% |
| Average Return on Invested Capital | 12.40% |

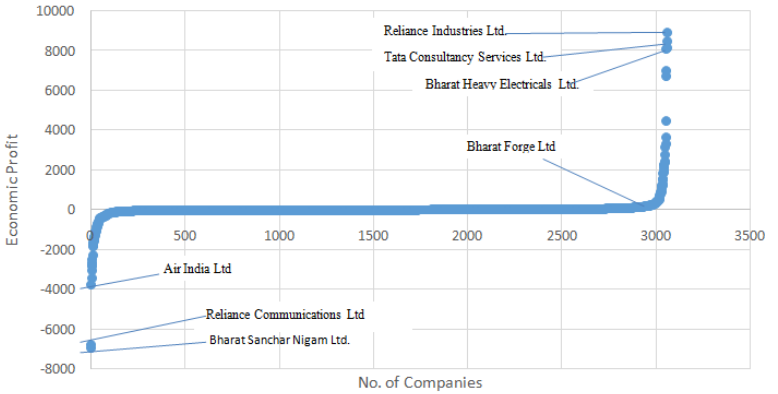


**Figure 1.** *Average Economic Profit 2011-2013, 3060 companies (Figures in Rs. Crore)*
**Source:** Authors' own construction.

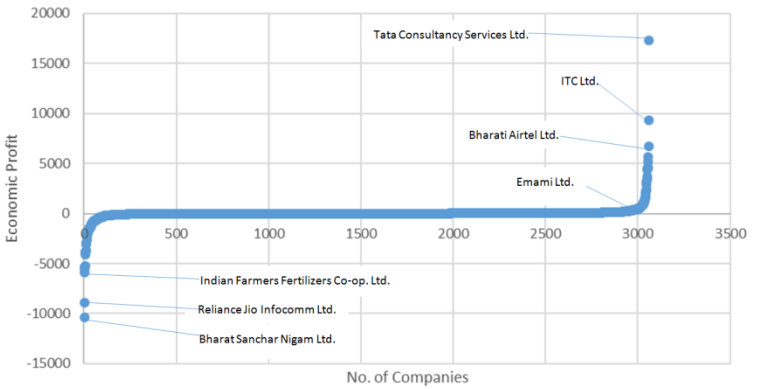


**Figure 2.** *Average Economic Profit 2014-2017, 3060 companies (Figures in Rs. Crore)*
**Source:** Authors' own construction.

Figures 1 and 2 are of the "Power Curve" as referred to in the book by Bradley *et al.*, (2018). Some companies earned much higher economic profit than the rest. Some companies earned negative economic profit. Most of the companies have earned little economic profit. These are profitable companies, but have not generated significant economic profit. Tata Consultancy Services (TCS) has come out to be a

high economic profit generator in both the periods, whereas Reliance Industries has lost its sheen in 2014-17. Bharti Airtel and ITC has shown significant improvement during the period and so has Bharat Forge. BSNL and Reliance Communications (later Reliance Jio Infocom) has done poorly, and we will observe later that they have adversely affected the overall performance of the sector.

For a better idea about the distribution of economic profit across companies, we divided the companies into quintiles in terms of economic profit. This is given in Tables 3 and 4.

**Table 3.** *Economic Profit, Quintile wise, for 3060 companies for year 2011-2013*
*(Figures in Rs. Crore)*

| | |
|---|---|
| Min | -6938.65 |
| 20th | -3.88 |
| 40th | 0.70 |
| 60th | 7.18 |
| 80th | 26.00 |
| Max | 8926.89 |

**Table 4.** *Economic Profit, Quintile wise, for 3060 companies for year 2014-2017*
*(Figures in Rs. Crore)*

| | |
|---|---|
| Min | -10406.12 |
| 20th | -7.21 |
| 40th | 0.34 |
| 60th | 6.87 |
| 80th | 28.74 |
| Max | 17326.44 |

Tables 3 and 4 suggest that in the lowest 20% quintile, the companies have done worse in terms of economic profit over the two time periods. However, from the 40$^{th}$ quintile upwards, the average economic profit has improved in 2014-17 over 2011-13. In the highest quintile, the average economic has improved significantly.

We now investigate whether size of a company affects economic profit. Figures 3 and 4 present the relationship for the highest and lowest quintile for the period 2014-17. Figure 3 shows that there is a positive relationship between capital employed and economic profit. That is, large companies have been able to generate higher economic profit than smaller companies in the highest quintile. However, companies like TCS and Vodafone have generated higher economic profit with lower invested capital than Tata Steel and Power Grid Corporation. The former companies are service providers and can generate higher returns with lower capital. Manufacturing will not have this edge and their returns will be lower.

Interestingly, for companies in the lowest quintile as shown in Figure 4, the effect is the reverse. With increase in invested capital, economic profit has a tendency to decrease. This has interesting implications for strategy formulation. Companies grow over time and this is through a process of capital accumulation.

However, companies in their effort to grow, at the end hurt themselves. It is possible, that there is an optimum scale. If this threshold cannot be crossed, economic profit generation may be difficult.

Further insight in the matter can be had from Figure 5. This shows the relationship between economic profit and invested capital for companies in the 40$^{th}$ to the 60$^{th}$ quintile in terms of economic profit. The diagram doesn't indicate any pattern. Many companies have been able to generate economic profit, irrespective of the size of capital invested.

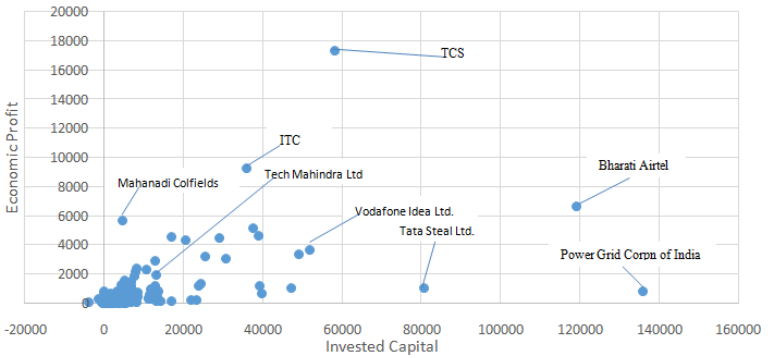


**Figure 3.** *Relationship between invested capital and economic profit for the highest quintile for 2014-2017 (Figures in Rs. Crore).*

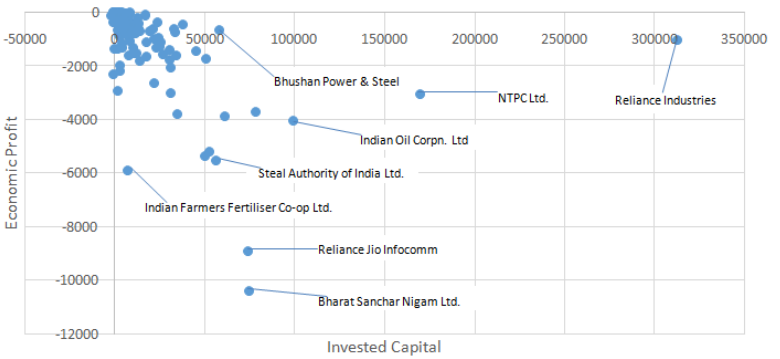


**Figure 4**. *Relationship between invested capital and economic profit for the lowest quintile for 2014-2017 (Figures in Rs. Crore)*

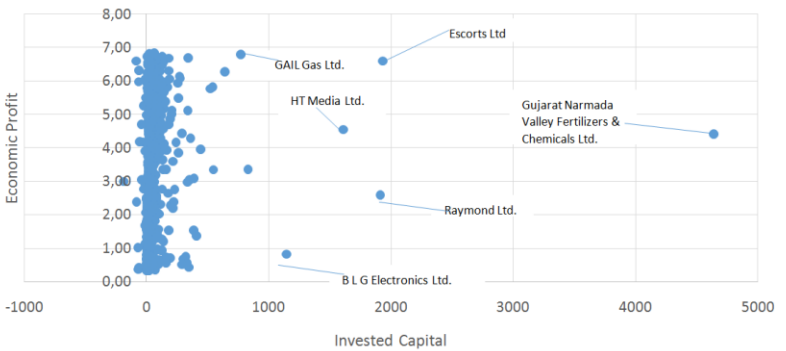


**Figure 5.** *Relationship between invested capital and economic profit for the 40th to the 60th quintile for 2014-2017*

Similar observations can be made for the period 2011-13 and these are presented in Figures 6 to 8.

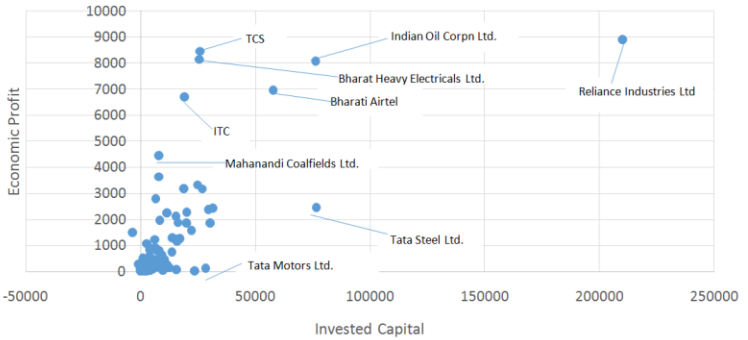


**Figure 6.** *Relationship between invested capital and economic profit for the highest quintile for 2011-2013 (Figures in Rs. Crore)*

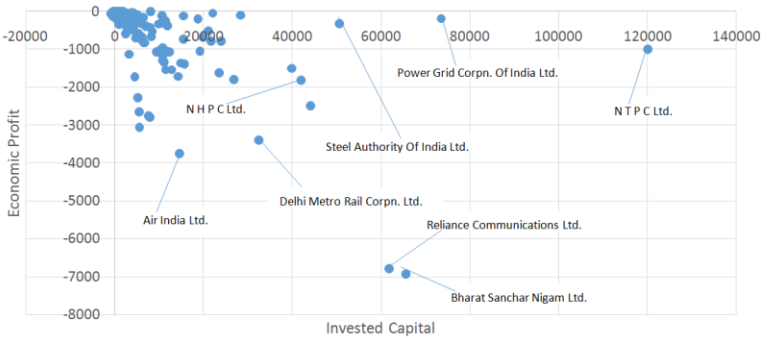


**Figure 7.** *Relationship between invested capital and economic profit for the lowest quintile for 2011-2013 (Figures in Rs. Crore)*

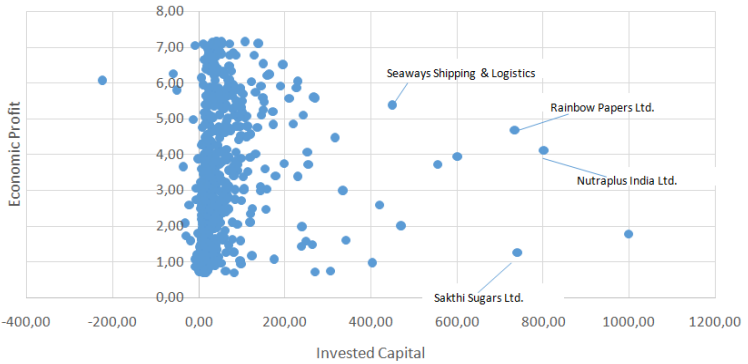


**Figure 8**. *Relationship between invested capital and economic profit for the 40th to the 60th quintile for 2011-2013*

The previous tables were constructed to understand the effect of size on economic profit where size was measured in terms of invested capital. We now look at size in terms of market capitalization of companies, and instead of quintiles, we focus on three classes of companies, viz, BSE large cap, BSE mid cap and BSE small cap. As before, we have considered only non-finance companies. Figures 9 to 14 show the "Power Curve" of these classes of companies for the years 2011-13 and 2014-17.

The figures suggest that number of companies generating negative economic profit are more in the small cap segment, as compared to the mid cap and large companies. This is expected, as these companies would be facing a strong competitive environment and are not that innovative to create a niche for themselves.

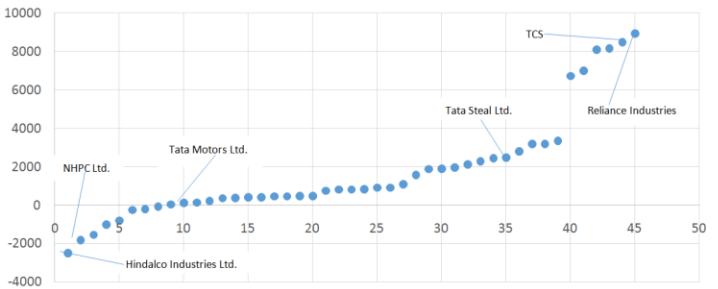


**Figure 9.** *Average Economic Profit 2011-2013, 45 companies, Large Cap (Figures in Rs. Crore)*

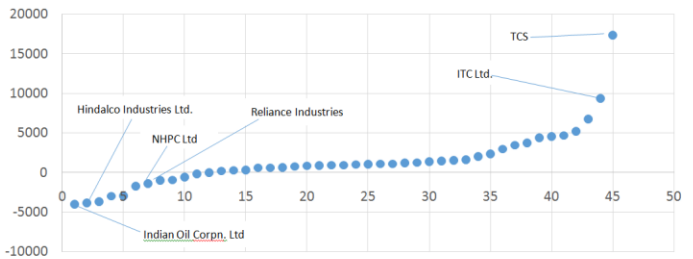


**Figure 10.** *Average Economic Profit 2014-2017, 45 companies, Large Cap (Figures in Rs. Crore)*

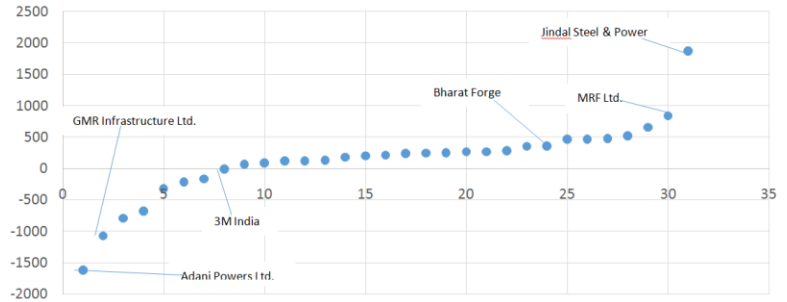


**Figure 11.** *Average Economic Profit 2011-2013, 31 companies, Mid Cap (Figures in Rs. Crore)*

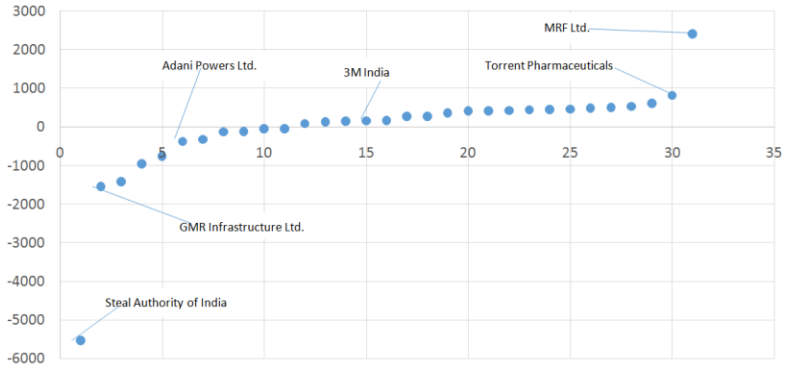


**Figure 12.** *Average Economic Profit 2014-2017, 31 companies, Mid Cap (Figures in Rs. Crore)*

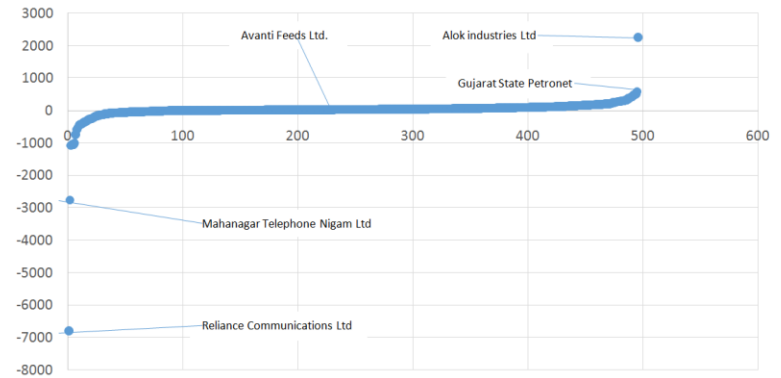


**Figure 13.** *Average Economic Profit 2011-2013, 496 companies, Small Cap (Figures in Rs. Crore)*

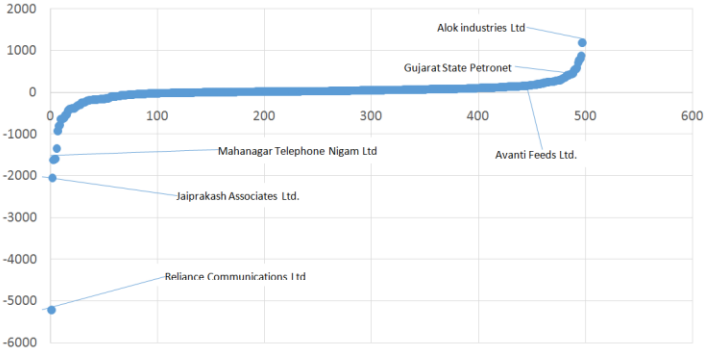


**Figure 14.** *Average Economic Profit 2014-2017, 496 companies, Small Cap (Figures in Rs. Crore)*

Table 5 and Figure 15 show that, for the companies considered under different market capitalization, there has been a deterioration in average economic profit, calculated over companies and time period, for the period 2014-17 over 2011-13, for large cap, mid cap and small cap companies.

**Table 5.** *Change in average economic profit over two time periods, cap wise (Figures in Rs. Crore)*

| Cap | 2011-13 | 2014-17 |
|---|---|---|
| Small Cap | 10.64 | 6.10 |
| Mid Cap | 120.82 | -54.61 |
| Large Cap | 1741.24 | 1363.23 |

**Source:** Authors' own construction

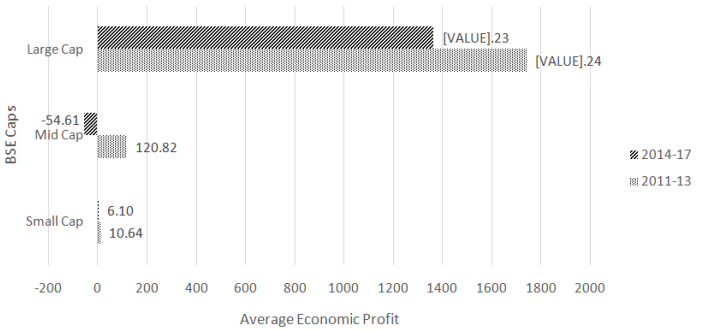


**Figure 15.** *Average economic profit over two time periods, cap wise (Figures in Rs. Crore)*

In the beginning, we had observed that the literature has indicated the importance of the industry for company performance. This was effectively highlighted by Bradley *et al.*, (2018), who emphasized that investment strategy should be guided by relative position of the industry, rather than inward looking plans for growth. Figures 16 and 17 present the industry-wide distribution of average economic profit of the various sectors for the years 2011-13 and 2014-17 respectively.

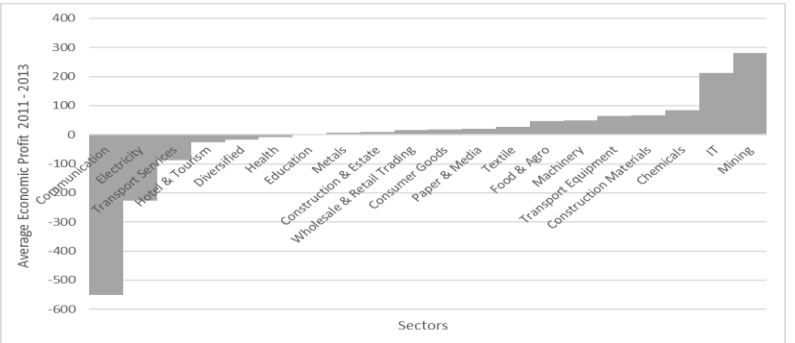


**Figure 16.** *Sector-wise average economic profit for the period 2011-13 (Figures in Rs. Crore)*

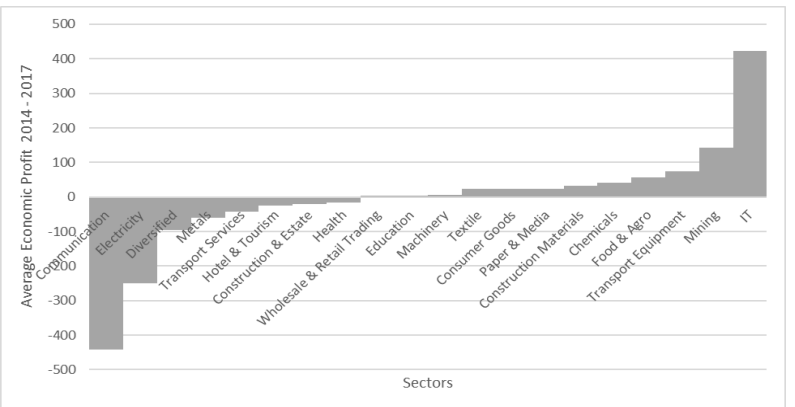


**Figure 17.** *Sector-wise average economic profit for the period 2014-17 (Figures in Rs. Crore)*

It may be observed that sectors like IT, Mining, Chemicals and Consumer Goods have generated economic profit consistent over the period. However, sectors like Communication, Electricity, Transport Services, Hotels and Tourism have not generated positive economic profit. The machinery sector has lost its shine in 2014-17 as compared to 2011-13.

It clearly emerges from the figures that certain sectors have generated negative profit, on an average, throughout the period 2011-17. Any amount of inward looking strategy formulation may not work for companies in these categories. Further, it also reflects the economic environment and external factors that have shaped the fortune of these sectors. There has also been a change in relative position of the sectors.

Table 6 and Figure 18 show sector-wise average economic profit over two time periods. These indicate the relative performance of the sectors over the two time periods. Table 7 and Figure 19 provides information on sector-wise average rate of profit over two time periods.

**Table 6.** *Sector-wise average economic profit over two time periods (Figures in Rs. Crore)*

| Sectors | 2011-2013 | 2014-2017 |
|---|---|---|
| Communication | -551.13 | -442.50 |
| Electricity | -226.40 | -249.87 |
| Transport Services | -88.11 | -42.15 |
| Hotel & Tourism | -25.76 | -25.83 |
| Diversified | -17.33 | -94.45 |
| Health | -7.19 | -15.40 |
| Education | 0.64 | 3.13 |
| Metals | 7.08 | -59.93 |
| Construction & Estate | 8.63 | -21.60 |
| Wholesale & Retail Trading | 17.04 | 2.76 |
| Consumer Goods | 17.73 | 24.25 |
| Paper & Media | 19.41 | 24.64 |
| Textile | 27.02 | 23.40 |
| Food & Agro | 46.43 | 55.92 |

| | | |
|---|---|---|
| Machinery | 49.29 | 6.43 |
| Transport Equipment | 64.79 | 75.26 |
| Construction Materials | 66.46 | 32.73 |
| Chemicals | 84.63 | 41.41 |
| IT | 212.70 | 423.60 |
| Mining | 281.24 | 142.50 |

**Source:** Authors' own construction.

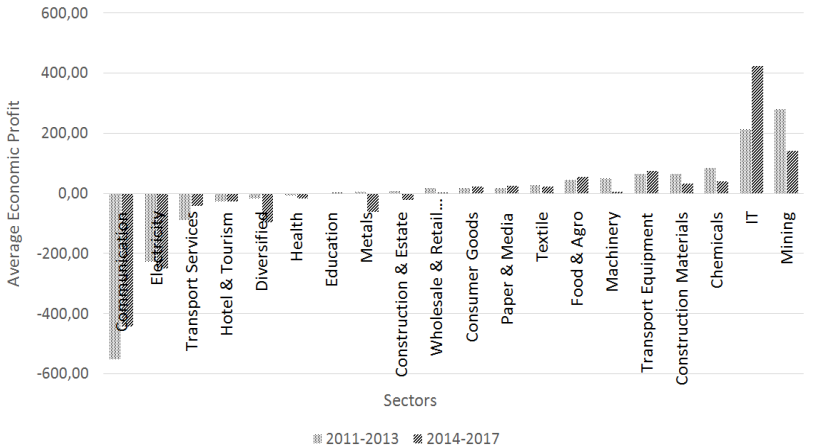


**Figure 18.** *Sector-wise average economic profit over two time periods (Figures in Rs. Crore)*

**Table 7.** *Sector-wise average rate of profit over two time periods*

| Sectors | 2011-2013 | 2014-2017 |
|---|---|---|
| Transport Service | 5.49% | 9.45% |
| Hotel & Tourism | 5.90% | 6.21% |
| Communication | 6.95% | 9.57% |
| Electricity | 8.50% | 9.28% |
| Health | 9.57% | 9.00% |
| Diversified | 10.28% | 4.53% |
| Metals | 12.50% | 8.93% |
| Construction & Estate | 12.93% | 10.16% |
| Education | 13.05% | 18.76% |
| Wholesale & Retail | 14.91% | 12.34% |
| Construction Material | 17.73% | 14.15% |
| Paper & Media | 18.53% | 18.87% |
| Chemical | 19.22% | 14.39% |
| Textile | 20.24% | 18.07% |
| Transport Equipment | 20.59% | 19.03% |

| | | |
|---|---|---|
| Machinery | 25.56% | 13.60% |
| Food & Agro | 26.25% | 19.10% |
| Consumer Goods | 27.20% | 25.56% |
| IT | 29.92% | 31.12% |
| Mining | 31.35% | 17.64% |

**Source:** Authors' own construction.

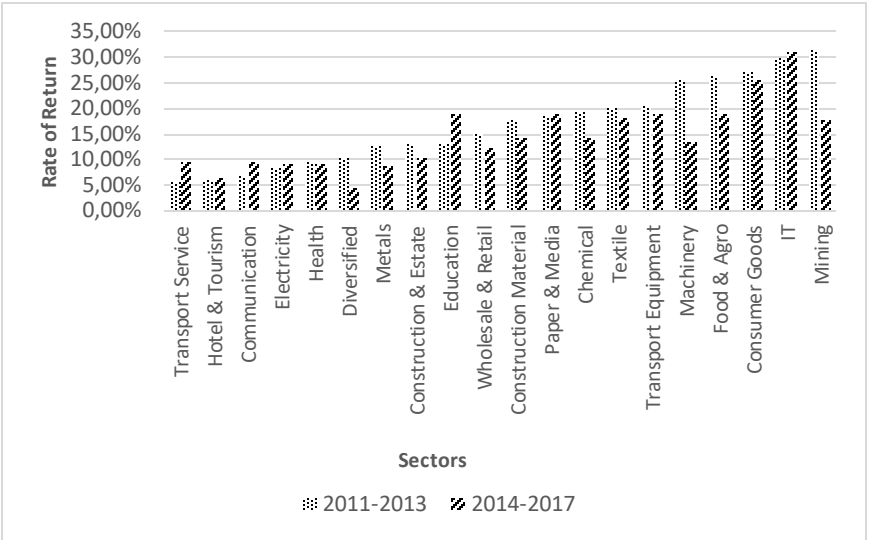


**Figure 19.** *Sector-wise average rate of profit over two time periods*
**Source:** Authors' own construction

We then investigated the frequency distribution of average economic profit over the two time periods to ascertain in which profit range most companies lie, irrespective of the sector and market capitalization. We found that almost 85% to 87% of the companies were in the economic profit range of Rs. - 81.12 crore to Rs.93.88 crore for the years 2011-13 and 2014-17 respectively.

For better insight, we further considered the frequency distribution of economic profit for companies in the economic profit range of Rs. - 81.12 crore to Rs.93.88 crore. These are presented in Tables 8 and 9 for the years 2011-13 and 2014-17 respectively. The tables show that most companies within this group earned marginal loss to

marginal profit of Rs. - 1.05 crore to Rs. 8.945 crore in both the years. In the overall scheme of things, in our sample, around 33 to 35 percent of companies earned positive economic profit during the time period.

**Table 8.** *Frequency distribution of companies in the economic profit range of Rs. - 81.12 crore to Rs.93.88 crore in 2011-13*

| Range in Rs. Crore | Frequency |
|---|---|
| -81.054 to -71.054 | 15 |
| -71.054 to -61.054 | 20 |
| -61.054 to -51.054 | 17 |
| -51.054 to -41.054 | 34 |
| -41.054 to -31.054 | 36 |
| -31.054 to -21.054 | 59 |
| -21.054 to -11.054 | 105 |
| -11.054 to -1.05 | 370 |
| -1.05 to 8.945 | 1121 |
| 8.94 to 18.946 | 351 |
| 18.946 to 28.946 | 194 |
| 28.946 to 38.946 | 113 |
| 38.946 to 48.946 | 73 |
| 48.946 to 58.946 | 45 |
| 58.946 to 68.946 | 43 |
| 68.946 to 78.946 | 32 |
| 78.946 to 88.946 | 26 |
| 88.946 to 98.946 | 13 |
| Grand Total | 2667 |

**Source:** Authors' own construction

**Table 9.** *Frequency distribution of companies in the economic profit range of Rs. - 81.12 crore to Rs.93.88 crore in 2014-17*

| Range in Rs. Crore | Frequency |
|---|---|
| -81.054 to -71.054 | 17 |
| -71.054 to -61.054 | 21 |
| -61.054 to -51.054 | 14 |
| -51.054 to -41.054 | 31 |
| -41.054 to -31.054 | 47 |
| -31.054 to -21.054 | 68 |
| -21.054 to -11.054 | 127 |
| -11.054 to -1.054 | 402 |
| -1.054 to 8.946 | 993 |

| | |
|---|---|
| 8.946 to 18.946 | 331 |
| 18.946 to 28.946 | 183 |
| 28.946 to 38.946 | 97 |
| 38.946 to 48.946 | 70 |
| 48.946 to 58.946 | 61 |
| 58.946 to 68.946 | 47 |
| 68.946 to 78.946 | 29 |
| 78.946 to 88.946 | 42 |
| 88.946 to 98.946 | 13 |
| Grand Total | 2593 |

**Source:** Authors' own construction

## Concluding remarks

The chapter provides a framework to understand the performance of Indian over two time periods 2011-13 and 2014-17. Economic profit is the metric that has been considered for measuring performance. As defined in the paper, economic profit is the surplus left over of operating profit, after accounting for the opportunity cost of capital. Any company should at least earn its opportunity cost, or the hurdle rate.

The data collated for 3060 companies indicates that the average economic profit has gone down in 2014-17 from 2011-13. The average performance of large cap, mid cap and small cap companies have deteriorated. The "Power Curve" shows that significant number of companies barely made any economic profit in both time periods. Thus, all plans and strategies, have not yielded the required results. The industry level scenario indicates that companies have not generated economic profit, as many sectors have not generated average economic profit. Thus, industry level factors need to be seriously considered.

We also investigated the relationship between the level of economic profit and size as represented by invested capital. Here a quintile-wise exercise was performed. The data shows that scale necessarily does not play a role in generating economic profit. For the highest quintile, some

positive relationship can be seen. However, for the lowest quintile, a higher invested capital has resulted in lower economic profit. Thus, many investments have not yielded desirable results.

## References

Bradley, C., Hirt, M., & Smit, S. (2018). *Strategy Beyond the Hockey Stick*, John Wiley & Sons.

Collins, J., & Porras, J. (1994). *Built to Last - Successful Habits of Visionary Companies*, Harper Business.

Govindrajan, V. (2016). *The Three Box Solution - A Strategy for Leading Innovation*, Harvard Business Review Press.

Kim, W.C., & Mauborgne, R. (2005). *Blue Ocean Strategy*, Harvard Business School Publishing Corporation.

McGrath, R.G. (2013). *The End of Competitive Advantage*, Harvard Business Review Press.

Peters, T.J., & Waterman, R.H. (1982). *In Search of Excellence - Lessons from America's Best Run Companies*, Collins Business Essentials.

Porter, M. (1985). *Competitive Advantage – Creating and Sustaining Superior Performance*, Free Press.

Porter, M. (1986). What is strategy? *Harvard Business Review*.

**For Cited this Chapter:**

**Jhunjhunwala, A., Chaudhuri, T.D., & Bhamrah, G.K., (2020). Value creation by Indian companies: A comparative study over two time periods, *In* T.D. Chaudhuri (Ed.), *Overview of the Indian Economy: Empirical analysis of financial markets, exchange rate movements and industrial development* (pp.7-30), KSP Books: Istanbul.**

ISBN: 978-605-7736-93-2 (e-Book)
KSP Books 2020
© KSP Books 2020

**Copyrights**

Copyright for this Book is retained by the author(s), with first publication rights granted to the Book. This is an open-access Book distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by-nc/4.0 ).

# 2 Can portfolio returns exceed market return? An examination of the efficient market hypothesis for the Indian stock market

*By*
**Tamal Datta CHAUDHURI**
**& Gulshan Kaur BHAMRAH**

## Introduction

The financial literature is replete with attempts in predicting stock prices. In contrast to the Efficient Market Hypothesis, researchers have identified various factors that can influence stock returns and hence have used them for prediction purposes. The quality of results has varied, but the efforts continued. Going back to Graham & Dodd (1934) where they disregarded the fact that "good stocks (or blue chips) were sound investments regardless of the price paid for them", they distinguished between speculation and investment, and consequently emphasized on factors like management quality, earnings, dividends, capital structure and interest cover. Their work focused on building a healthy portfolio and the characteristics of their constituents. Implicit in their work was the theme that it pays to be careful while choosing

stocks and there were fundamental factors that the investors should carefully consider.

There are a large number of financial sector players who use indicators of technical analysis to predict stock price movements. Their evaluation of overpriced or underpriced stocks is based on technical indicators like moving average, momentum, stochastics, percentage retracement, MACD, RSI etc. It is their belief that future stock prices can be predicted and accordingly portfolio can be constructed. As market prices contain all information, present and future, they base their decisions regarding buying and selling of stocks on patternsof their price movements.

The mutual fund industry sells funds to subscribers based on returns that they can deliver. Stock choice and market timing are two factors that they focus on, implying that they believe that proper construction of an equity portfolio can generate returns. Mutual fund schemes available in India like Systematic Investment Plan (SIP), where each month a specific amount is put into a specific fund, are publicized as instruments which can generate returns above market returns over a long period of time.

Best-selling books on stock picking advise on looking at fundamentals of companies. Books and papers on behavioral finance have identified human traits and their effects on stock price movements. Various econometric methods have been used for prediction through frameworks that incorporate volatility. In short, relentless attempts are being made to demonstrate that money can be made in the stock market, implying thereby that stock markets are inefficient

The purpose of this chapter is to examine whether, based on certain parameters, it is possible to construct portfolio of stocks that can beat market returns. Accordingly, the plan of the chapter is as follows. A brief literature survey is presented in Section 2. The methodology for the study is

presented in Section 3. Section 4 presents the data and the results. Section 5 concludes the chapter.

## Literature review

Black & Scholes (1974) examine the effect of dividend policy on expected returns. They state that companies that declare higher dividend observe an increase in prices, and companies that reduce dividends face a temporary decline in prices. Investors find it difficult to understand that companies declare lower dividends to conserve funds for expansion purposes. They test the effect of dividend yield on expected return by considering a modified version of the CAPM model with dividend yield as an additional explanatory variable.

Basu (1977, 1983) analyzes the relationship between earnings yield, size and returns from common stocks and find that stocks with higher earnings to price, earned higher returns than those with low earnings to price. This result they obtained, even when the experiment was conducted for firms of various sizes.

Banz (1981) explores the relationship between size, as represented by market capitalization, and market yields. The study found that smaller firms have had higher risk adjusted returns, on average, than larger firms. Chan *et al.*, (1991) tried to explain returns from Japanese stocks in terms of earnings yield, size, book to market ratio, and cash flow yield and found significant relationship between these variables.

Fama & French (1988) explore the relationship between dividend yields and expected stock returns. Considering the dividend discount model, which has expected returns and discount factors involved, they show that dividend yields explains less than five per cent of variances in returns. As the time increases, the effect of dividend yields increases, and this is due to high autocorrelation of expected returns.

Jaffe, Kiem & Westerfield (1989) examine the effect of size and earnings to price ratio on stock returns. Their study focusses on the effect of the month of January, compared to other months. Further, they focus on small firms and turnaround firms.

Fama & French (1995) try to relate earnings from stocks with ratio of book value to market price (BE/ME) and size. Their contention is that low BE/ME stocks generate higher returns than high BE/ME stocks. Further, high BE/ME stocks signal poor earnings and low BE/ME stocks signal strong earnings.

Campbell & Shiller (1998) investigate whether stock prices can drift away from fundamentals like dividends or earnings for long periods, and whether there would be a tendency for prices to correct so as to keep the relationship between them at normal levels. In relation to the efficient market hypothesis, they test whether dividend price ratio or any other valuation ratio has the ability to predict future stock price movements.

Senyigit &Ag (2014) examine the effect of three independent variables namely P/E ratio (price to earnings ratio), P/B ratio (price to book ratio), and D/E ratio (debt to equity ratio) on stock returns in Turkey. For their sample and the time period under consideration, they did not find any significant relationship between the variables.

For further references, the reader can refer to Datta Chaudhuri, Ghosh & Eram (2016) where for prediction of stock returns, the explanatory variables (inputs/features) considered are Price-Earnings (P/E) Ratio, Price to Book Value (P/BV) Ratio, Debt Equity Ratio (DER), Interest Coverage Ratio (ICR), Gross Profit Margin (GPM), Dividend Pay-Out Ratio (DPR), Extent of Promoter Holding of Shares (EPH), Sectoral Returns (SR) and Volume (V). The chapter uses variables that have been tried in the literature, but in a

machine learning framework, where no linearity assumption is made and where there is continuous learning.

## Methodology

For portfolio formation, the attributes of companies that we focus in this chapter are Size as represented by Sales and Market Capitalization, P/E ratio, Market to Book value, Net Profit Margin, Dividend Yield and PEG ratio. The following presents how we have constructed the different portfolios, on the basis of the attributes.

a. The data is for the period 2013 to 2018

b. We have considered only manufacturing companies that are listed in the Indian stock exchanges, have made profits in all the years mentioned above, and have declared dividends

c. We have then ranked these companies by sales, to represent size, from smallest to largest, and have only focused on companies of sales of Rs.100 crore or more.

d. From this subset of companies, we have chosen the top 100 companies and the bottom 100 companies.

e. For each of these 100 companies, we have ranked them by the attributes and considered the lowest decile and the highest decile by the attributes.

f. We have considered these as our portfolios, and compared their returns with market returns.

Once the data is in place, we have estimated equation 1 for various portfolios constructed on the basis of the parameters mentioned above.

$$(r_p - r_{f)} = \alpha + \beta\ (r_m - r_f)$$

where $r_p$ is the returns from the portfolio, $r_f$ is the risk free rate, $r_m$ is market returns, $\alpha$ represents excess returns, and $\beta$ is the measure of systematic risk. If for any of the portfolios, the relationship generates a positive significant value of $\alpha$,

we can say that that the portfolio so constructed has been able to generate returns above the market.

We also perform the same exercise for portfolios constructed on the basis size represented by market capitalization. We perform separate exercises for large cap, mid cap and small cap companies as classified by the Bombay Stock Exchange. In this case, some banks were selected in the some of the portfolios.

For the regression, we collected daily data from 2014 to 2018 on market prices of the stocks in the respective portfolios. We computed monthly returns from the stocks and also monthly returns of the market index, NIFTY. For portfolio returns, we took the weighted average of the returns of the stocks in the portfolio, where each stock was given the same weight.

## Results

The results are presented in Tables 1, 2, 3 and 4.

a. Table 1 suggests that when the companies are ranked in increasing order of sales, if we consider the largest 100 companies, then rank them as per the PEG ratio, the P/E ratio, P/B ratio, NPM, and Dividend Yield, the respective portfolios consisting of first and last decile of these companies, have not generated returns greater than the market returns. From the regression, we did not get a significant value of $\alpha$. These companies feature in most mutual funds, most of them are part of NIFTY representing market sentiment, and are also mostly sought after by individual investors. They are stable companies with established products. For such set of companies to generate above market returns would has been difficult.

**Table 1.** *Regression Results for Top 100 Companies in terms of Sales*

| Parameter | Estimated $\alpha$ | Estimated $\beta$ | $R^2$ |
|---|---|---|---|
| PEG ratio | | | |
| First decile | 0.151248 | 0.967066 | 0.237544 |
| | (1.489683) | (4.250874) | |
| Last decile | 0.017138 | 1.027406 | 0.495642 |
| | (0.282177) | (7.549687) | |
| PE ratio | | | |
| First decile | 0.106188 | 1.140572 | 0.289429 |
| | (1.01395) | (4.860502) | |
| Last decile | 0.003414 | 0.78474 | 0.493945 |
| | (0.073346) | (7.524105) | |
| PB ratio | | | |
| First decile | 0.142375 | 1.116254 | 0.313144 |
| | (1.469626) | (5.142245) | |
| Last decile | 0.020405 | 0.599794 | 0.372116 |
| | (0.446924) | (5.862917) | |
| NPM | | | |
| First decile | 0.082961 | 1.334974 | 0.326393 |
| | (0.738186) | (5.301287) | |
| Last decile | -0.01134 | 0.57334 | 0.336908 |
| | (-0.24057) | (5.428539) | |
| DIVIDEND YIELD | | | |
| First decile | 0.04522 | 0.659885 | 0.229433 |
| | (0.638083) | (4.155626) | |
| Last decile | 0.109834 | 1.135464 | 0.328866 |
| | (1.155483) | (5.331129) | |

**Notes.** Figures in brackets indicate t-values.
**Source.** Authors' own construction

b. It is of interest to note from Table 2 that of the stocks in the bottom 100 companies, when ranked in terms of the different parameters and two portfolios formed consisting of the first and last decile of companies, these portfolios havegenerated more than market returns over the period 2013 to 2018. When PEG ratio is considered, the regression results for both the first and last decile of companies have generated positive significant $\alpha$. This is also true for dividend yield. For P/E ratio, P/B ratio and NPM, regression for the first decile of companies has generated positive

significant $\alpha$. We can infer that portfolio of smaller companies have been able to generate returns higher than the market.

**Table 2.** *Regression Results for Bottom 100 companies in terms of Sales*

| Parameter | Estimated $\alpha$ | Estimated $\beta$ | $R^2$ |
|---|---|---|---|
| PEG ratio | | | |
| First decile | 0.20915 | 1.142582 | 0.273839 |
| | (1.918221)** | (4.676756) | |
| Last decile | 0.16469 | 0.829949 | 0.227309 |
| | (1.836612)** | (4.130663) | |
| PE ratio | | | |
| First decile | 0.179566 | 0.927386 | 0.200806 |
| | (1.656237)* | (3.817474) | |
| Last decile | 0.121678 | 0.517796 | 0.093536 |
| | (1.288141) | (2.446403) | |
| PB ratio | | | |
| First decile | 0.208145 | 0.769389 | 0.112695 |
| | (1.645253)* | (2.714126) | |
| Last decile | 0.07404 | 0.613396 | 0.196014 |
| | (1.017044) | (3.760391) | |
| NPM | | | |
| First decile | 0.216685 | 0.781017 | 0.1162 |
| | (1.716683)** | (2.761467) | |
| Last decile | 0.051593 | 0.533153 | 0.118337 |
| | (0.604985) | (2.790124) | |
| Dividend Yield | | | |
| First decile | 0.129733 | 0.570196 | 0.161085 |
| | (1.701344)** | (3.337199) | |
| Last decile | 0.173241 | 0.674023 | 0.12299 |
| | (1.642499)* | (2.85198) | |

**Notes.** Figures in brackets indicate t-values. **Source.** Authors' own construction

**Table 3.** *Regression Results for Market Leaders*

| Parameter | Estimated $\alpha$ | Estimated $\beta$ | $R^2$ |
|---|---|---|---|
| SECTOR Leaders | | | |
| First quintile | 0.117462 | 0.839067 | 0.326171 |
| | (1.662055)* | (5.29861) | |
| Last quintile | 0.067977 | 0.813906 | 0.451836 |
| | (1.293962) | (6.914312) | |

**Notes:** Figures in brackets indicate t-values. **Source:** Authors' own construction

c. The portfolios for Table 3 were constructed by taking market leaders, in terms of sales, from various sectors in India, and then forming two portfolios with the lowest sales and highest sales. These portfolios have sectoral diversification and the companies are leaders in their fields. Interestingly, we observe that the lowest quintile companies have been able to generate above market returns.

d. Since our focus was not on systematic risk, we have not made any comment on Estimated $\beta$. Observation will show that they are in each case statistically significant.

e. As we move from highest 100 companies to lowest 100 companies the values of $R^2$ declined. The association with the market fell. However, for market leaders, the $R^2$ was the highest.

f. In all the above exercises, the companies were initially ranked by sales, and then portfolios were constructed on the basis of different parameters. In Table 4 we report regression results for companies based on market capitalization. We consider the companies that are listed in the Bombay Stock Exchange (BSE) Large Cap Index, the BSE Mid Cap Index and the BSE Small Cap Index. In each of these indices we take the first and the last quintile of companies and form six portfolios, ranked by the parameters E/P, Dividend Yield and Net Profit Margin (NPM). Our regression results show that portfolios formed on the basis of E/P could not generate positive significant $\alpha$. In terms of dividend yield, it is the portfolio of companies belonging to the highest quintile in the large cap sector that has generated greater than market returns. When portfolios are formed on the basis of NPM, the top quintile of all the mid cap and large cap companies have generated significant positive $\alpha$. The value of $\alpha$ for the top decile in the small cap segment is marginally significant. The results indicate that business efficiency matters for portfolio choice.

**Table 4.** *Regression Results for Portfolios based on Market Capitalization*

| Market capitalization wise companies | Estimated $\alpha$ | Estimated $\beta$ | $R^2$ |
|---|---|---|---|
| E/P | | | |
| Small cap | | | |
| Top | -0.05844<br>(-0.498) | 1.2465<br>(4.704) | 0.2761 |
| Bottom | 0.032105<br>(0.326525) | 1.092378<br>(4.92067) | 0.2945 |
| Mid cap | | | |
| Top | 0.016657<br>(0.192623) | 1.191129<br>(6.1007) | 0.390879 |
| Bottom | -0.01325<br>(-0.1756) | 0.99878<br>(5.8637) | 0.372182 |
| Large Cap | | | |
| Top | -0.00406<br>(-0.0574) | 1.010083<br>(6.3254) | 0.40823 |
| Bottom | -0.00281<br>(-0.03948) | 0.869688<br>(5.4041) | 0.334895 |
| DIVIDEND YIELD | | | |
| Small cap | | | |
| Top | -0.03154<br>(-0.32308) | 1.687465<br>(7.310929) | 0.479586 |
| Bottom | 0.058154<br>(0.848977) | 1.271123<br>(7.84969) | 0.515122 |
| Mid cap | | | |
| Top | -0.00742<br>(-0.08809) | 1.631825<br>(8.19673) | 0.536691 |
| Bottom | 0.014007<br>(0.296022) | 1.024492<br>(9.158727) | 0.59121 |
| Large Cap | | | |
| Top | 0.085126<br>(1.947218)** | 1.295771<br>(12.53819) | 0.730491 |
| Bottom | 0.028667<br>(0.619401) | 0.931945<br>(8.517795) | 0.555735 |
| NPM | | | |
| Small cap | | | |
| Top | 0.157043<br>(1.597455)* | 1.601625<br>(6.891641) | 0.450209 |
| Bottom | 0.085033<br>(1.205068) | 1.26634<br>(7.59145) | 0.498401 |

| | | | |
|---|---|---|---|
| Mid cap | | | |
| Top | 0.137844<br>(2.477655)** | 1.161103<br>(8.828252) | 0.573335 |
| Bottom | 0.061681<br>(1.307419) | 1.04918<br>(9.407344) | 0.60409 |
| Large Cap | | | |
| Top | 0.135057<br>(3.508869)** | 1.085057<br>(11.92486) | 0.710293 |
| Bottom | 0.004665<br>(0.105233) | 0.940477<br>(8.973351) | 0.581291 |

**Notes.** Figures in brackets indicate t-values. **Source.** Authors' own construction

The list of companies belonging to the portfolios discussed above, are available from the authors on request.

## Conclusion

The objective of the chapter was to explore whether portfolios of stocks could be constructed which would deliver returns higher than the market. The basis of constructing the portfolios were size, Price/Earnings ratio, Market Price/Book Value ratio, Net Profit Margin, Dividend Yield, PEG ratio and Earnings to Price ratio. Our portfolios were constructed in the year 2013, and their performance was evaluated over a fiveyear period from 2014 to 2018. Our results show that for the largest companies in terms of sales, the portfolios could not generate above market returns for any of the parameters. Whereas, for the smallest of the companies, some of the portfolios could deliver excess returns.

It is interest to note that portfolios formed with market leaders in different industries, could deliver excess returns. Thus, leadership needs to be taken into consideration for portfolio formation.

When we controlled for market capitalization to represent size, we found that portfolios formed from largest cap companies on the basis of dividend yield or net profit margin, generated above market returns. Thus, companies

with highest market capitalization, that are liquid, efficient and dividend paying are preferred by investors.

Overall, we laid out certain principles for portfolio formation, and our results are specific for a certain period of time. We observed that certain portfolios could generate above market returns, whereas some didn't. It would be our endeavor now to examine on what basis various equity mutual funds in India choose their portfolio and whether focused funds like dividend yield funds are better than funds based on market capitalization or on size.

## References


Black, F.,& Scholes M (1974). The effects of dividend yield and dividend policy on common stock prices and returns, *Journal of Financial Economics*, 1(1), 1-22.doi. 10.1016/0304-405X(74)90006-3

Basu, S. (1977). Investment performance of common stocks in relation to their price earnings ratios: A test of the efficient market hypothesis, *Journal of Finance*, 32, 663-682. doi. 10.1111/j.1540-6261.1977.tb01979.x

Banz, R.W. (1981). The Relationship between Return and Market Value of Common Stocks, *Journal of Financial Economics*, 9(1), 3-18. doi. 10.1016/0304-405X(81)90018-0

Basu, S. (1983). The relationship between earnings yield, market value and return for NYSE common stocks: Further evidence, *Journal of Financial Economics*, 12(1), 129-156. doi. 10.1016/0304-405X(83)90031-4

Campbell, J.Y., & Shiller, R.J. (1998). Valuation ratios and the long-run stock market outlook, *The Journal of Portfolio Management*, 24(2), 11-26. doi. 10.3905/jpm.24.2.11

Chan, L., Hamao, Y., & Lakonishok, J. (1991). Fundamentals and stock returns in Japan, *The Journal of Finance*, 46(5), 1739-1764. doi. 10.1111/j.1540-6261.1991.tb04642.x

Datta Chaudhuri, T., Ghosh, I., & Eram, S. (2016). Application of unsupervised feature selection, machine learning and evolutionary algorithm in predicting stock returns – A study of Indian firms, *The IUP Journal of Financial Risk Management*, 13(3), 20-46.

Database CMIE Prowess, (2019). [Retrieved from]., [Retrieved from]., [Retrieved from].

Fama, E.F., & French, K.R. (1995). Size and book-to-market factors in earnings and returns, *Journal of Finance*, 50(1), 131-155. doi. 10.1111/j.1540-6261.1995.tb05169.x

Fama, E.,& French, K.R. (1988). Dividend yields and expected stock returns, *Journal of Financial Economics*, 22(1), 3-25. doi. 10.1016/0304-405X(88)90020-7

Graham, B., & Dodd, D. (1934). *Security Analysis*, Whittlesey House, McGraw hill Book Co.

Jaffe, J., Keim, D.B.,& Westerfield, R. (1989). Earnings yields, market values, and stock returns, *Journal of Finance*, 44(1), 135-148. doi. 10.1111/j.1540-6261.1989.tb02408.x

Senyigit, Y.B.,& Ag, Y. (2014). Explaining the cross section of stock returns: A comparative study of the United States and Turkey, *Procedia - Social and Behavioral Sciences* 109, 327-332. doi. 10.1016/j.sbspro.2013.12.466

**For Cited this Chapter:**

**Chaudhuri, T.D., & Bhamrah, G.K., (2020). Can portfolio returns exceed market return? An examination of the efficient market hypothesis for the Indian stock market, *In* T.D. Chaudhuri (Ed.), *Overview of the Indian Economy: Empirical analysis of financial markets, exchange rate movements and industrial development* (pp.31-44), KSP Books: Istanbul.**

ISBN: 978-605-7736-93-2 (e-Book)
KSP Books 2020
© KSP Books 2020

**Copyrights**

Copyright for this Book is retained by the author(s), with first publication rights granted to the Book. This is an open-access Book distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by-nc/4.0 ).

# 3 An alternative framework for time series decomposition and forecasting and its relevance for portfolio choice A comparative study of the Indian consumer durable and small cap sectors

*By*
Jaydip SEN
& Tamal Datta CHAUDHURI

## Introduction

Prediction of stock prices has been one of the biggest challenges to researchers, particularly to those belonging to the Artificial Intelligence (AI) community. Various technical, fundamental, and statistical indicators have been proposed and used with varying results. In our recent research work, we have proposed a new way of looking at portfolio diversification and prediction of stock returns (Sen & Datta Chaudhuri, 2016a; Sen & Datta Chaudhuri, 2016b). It has been postulated that different sectors in an economy do not behave uniformly, and sectors differ from each other in terms of their trend pattern, their seasonal characteristics and also their randomness. While the randomness aspect has been the cornerstone of *Efficient Market Hypothesis*, the literature trying to prove or disprove it, has delved into the various fundamental characteristics of each company and have come

up with different results. For example, Datta Chaudhuri, Ghosh and Eram applied Random Forest and Dynamic Evolving Neural-Fuzzy Inference System (DENFIS) to predict stock returns of mid cap Indian firms (Datta Chaudhuri *et al*, 2016). Our contention has been that, besides their fundamental characteristics, performances of companies depend on the performance of the sector to which they belong, and each sector has its own reason for growth or stagnation. The reasons behind the fortunes of the IT sector in India are different from those of the Steel sector or the Pharmaceutical sector, and these differences have to be factored in for portfolio choice and also churning of the portfolio.

In this work, we focus on the time series pattern of two sectors in India, namely the Consumer Durables sector and the Small Cap sector. We first demonstrate that the time series decomposition approach proposed provides us with deeper understanding of the behavior of a time series by observing the relative magnitudes of its three components namely trend, seasonal and random and also enables us to validate some hypotheses. For example, the Consumer Durables sector in India is known to display seasonal characteristics and the Small Cap sector in India is speculative in nature, and hence should have strong random components. The decomposition approach enables us to study the seasonal components and the random components of these two sectors separately and validate these hypotheses. With regard to the seasonal components, the decomposition approach also helps us to understand during which months which sectors are strong/weak so that buy/sell decisions about the stocks of companies in those sectors can be made effectively. The sectors with dominant random components in their time series, however, can be used for pure speculative gains.

Second, we propose an extensive framework for time series forecasting and a quantitative approach to analyze the change in behavior of the constituents (i.e., the trend, the seasonal, and the random component) of a time series over a long period of time. We have applied five techniques of forecasting using R environment and also provided a detailed guideline about which technique to use under what situations and for what type of time series behavior.

Third, we have presented a robust quantitative approach for analyzing any change in behavior of the constituents (i.e., the trend and the seasonal component) of a time series over a long period of time. If the behavior of the components of a time series does not change significantly over time, it is possible to design very robust forecasting framework for the time series.

The rest of the chapter is organized as follows. Section 2 briefly discusses the methodology in constructing various time series and decomposing the time series into its components. It also presents a brief outline on the forecasting frameworks designed in this work using the R programming language. Section 3 provides a detailed discussion on the methods of decomposition, the decomposition results of both sectors under study, and an analysis of the results. In addition, it presents two hypotheses and their validations using our experimental results. In Section 4, five robust forecasting techniques are proposed and a framework for analyzing the behavior of the structural constituents (i.e., the trend, the seasonal, and the random component) of a time series using the R programming environment. Section 5 presents detailed results of forecasting using all the methods that we proposed in Section 4. The forecasting methods are compared based on some suitably chosen metrics and a critical comparative analysis is presented for the proposed methods of forecasting. We have also analyzed the reason why certain

methods perform better compared to the others for certain time series under certain situations. In Section 6, we discuss some related work in the current literature. Finally, Section 7 concludes the chapter.

## Methodology

In this section, we provide a brief outline of the methodology that we have followed in our work. However, each of the following sections contains detailed discussion on the methodology followed in the work related to that Section. We have used the *R programming language* (Ihaka & Gentleman, 1996) for data management, data analysis and presentation of results. R is an open source language with very rich libraries that is ideally suited for data analysis work. In this work, we use daily data from the Bombay Stock Exchange (BSE) on BSE Consumer Durables Index and BSE Small Cap index for the period January 2010 to December 2015. The daily index values are first stored in two plain text files – each sector data in one file. The daily data are then aggregated into monthly averages resulting in 70 values in the time series data. These 70 monthly average values for each sector are stored into two different plain text files – each sector monthly average in one file. The records in the text file for each sector are read into an R variable using the *scan( )* function in R. The resultant R variable is converted into a monthly time series variable using the *ts( )* function defined in the *TTR* library in the *R* programming language. The monthly time series variable in R is now an aggregate of its three constituent components: (i) trend, (ii) seasonal, and (iii) random. We then decompose the time series into its three components. For this purpose, we use the *decompose( )* function defined in the *TTR* library in R. The decomposition results enable us to make a comparative analysis of the behavior of the two different time series belonging to two

different sectors. We validate two hypotheses by our deeper analysis of the decomposition results.

After a detailed analysis of the decomposition results, we enter into our second endeavor in this work. We have designed and analyzed five robust forecasting methods using the *HoltWinters( )* function, *Auto Regressive Integrated Moving Average* (ARIMA) framework, and an approach based on computation of the aggregate of the trend and seasonal components – all in the *R* computing framework. A detailed comparative analysis, highlighting which method performs best under what situation and for what type of time series, is also presented.

In our previous work, we have highlighted the effectiveness of time series decomposition approach for robust analysis and forecasting of the Indian Auto sector (Sen & Datta Chaudhuri, 2016a; Sen & Datta Chaudhuri, 2016b). In this work, we have compared two different sectors – Indian Consumer Durables sector and the Indian Small Cap sector and proposed guidelines and frameworks for comparing different sectors based on time series decomposition studies. Based on our analysis, we have also validated two hypotheses on the behavior of the two sectors under study. We have also analyzed and determined what forecasting technique to use based on the behavior of the time series and also have highlighted the reasons why some forecasting approaches perform better in comparison with other approaches under certain situations.

## Time series decomposition results

We now present the methods that we have followed to decompose time series for both BSE Consumer Durables Index and BSE Small Cap Index and then present the results that we have obtained from the decomposition work.

For both the sectors, we have first taken the daily index values from January 2010 to December 2015 and saved the

values in plain text (.txt) files. From these daily index values, we have computed the month averages and saved the monthly average values in two different text files. Each of these text files contained 72 values (6 years, each year containing 12 month average values). We used R language function *scan( )* to read these text files and store them into appropriate R variables. Then, we converted these R variables into time series variables using the R function *ts( )* defined in the package *TTR*. Once these time series variables are constructed, we have used the *plot( )* function in R to derive the displays of the time series. The time series for the Consumer Durables sector and the Small Cap sector are presented in Figure 1 and Figure 3 respectively.

The plots of the time series for the two sectors provide us an overall idea about how the two sectors have performed over the period under consideration (i.e., January 2010 – December 2015). Figure 2 and Figure 4 present the results of decomposition for the times series of the Consumer Durables sector and the Small Cap sector respectively. Each of these two figures has four boxes arranged in a stack. The boxes depict the overall time series, the trend, the seasonal and the random component respectively, arranged from top to bottom.

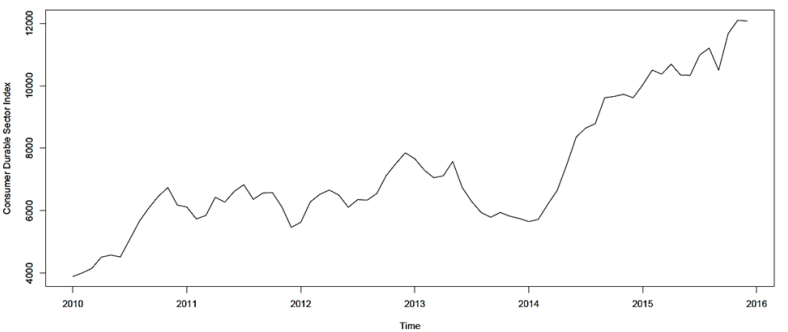


**Figure 1.** *The Consumer Durables Sector Index Time Series (Jan 2010 – Dec 2015)*

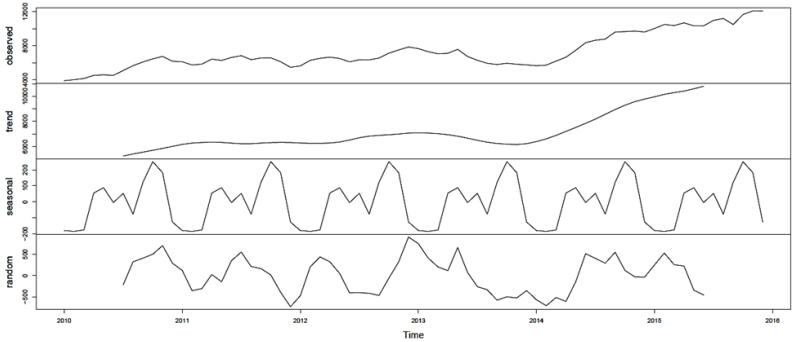


**Figure 2.** *Decomposition of Consumer Durables Sector Index Time Series into Trend, Seasonal and Random Components (Jan 2010 – Dec 2015)*

Table 1 and Table 2 present the numerical values of the time series data and its three components for the Consumer Durables sector and the Small Cap sector respectively. It may be interesting to observe that the values of the trend and the random components are not available for the period January 2010 – June 2010 and also for the period July 2015 – December 2015. Since the *decompose( )* function in R uses a 12 month moving average method for computing the trend component, in order to compute the trend value for January 2010, we need time series data from July 2009 to June 2010. However, since we have used time series data from January 2010 to December 2015, the first trend value the *decompose( )* function could compute was for the month of July 2010 and the last month being June 2014. For computing the seasonal component, the *decompose( )* function first *detrends* (subtracts the trend component from the overall time series) the time series and arranges the time series values in a 12 column format. The seasonal values for each month is derived by computing the averages of each column. The value of the seasonal component for a given month remains the same for the entire period under study. The random components are obtained after subtracting the sum of the corresponding

trend and seasonal components from the overall time series values. Since the trend values for the period January 2010 – June 2010 and July 2015 – December 2015 are missing, the random components for those periods could not be computed as well.

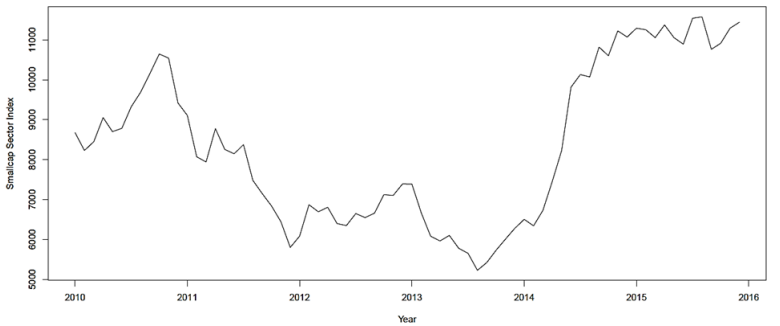


**Figure 3.** *The Small Cap Sector Index Time Series (Jan 2010 – Dec 2015)*

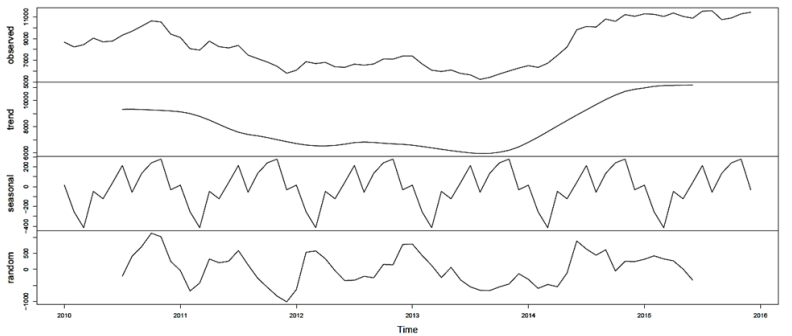


**Figure 4.** *Decomposition of Small Cap Sector Index Time Series into Trend, Seasonal and Random Components (Jan 2010 – Dec 2015)*

**Table 1.** *Consumer Durables Sector Aggregate Index and its Components (2010-2015)*

| Year | Month | Aggregate | Trend | Seasonal | Random |
|---|---|---|---|---|---|
| 2010 | Jan | 3890 | | -180 | |
| | Feb | 4006 | | -185 | |
| | Mar | 4150 | | -176 | |
| | Apr | 4512 | | 55 | |
| | May | 4575 | | 88 | |
| | Jun | 4518 | | -5 | |
| | Jul | 5084 | 5248 | 53 | -216 |

| | | | | | |
|---|---|---|---|---|---|
| | Aug | 5658 | 5412 | -78 | 323 |
| | Sep | 6086 | 5555 | 122 | 410 |
| | Oct | 6458 | 5705 | 251 | 501 |
| | Nov | 6742 | 5856 | 182 | 704 |
| | Dec | 6179 | 6014 | -127 | 291 |
| 2011 | Jan | 6115 | 6175 | -180 | 120 |
| | Feb | 5735 | 6277 | -185 | -357 |
| | Mar | 5844 | 6326 | -176 | -306 |
| | Apr | 6429 | 6351 | 55 | 24 |
| | May | 6271 | 6330 | 88 | -147 |
| | Jun | 6620 | 6275 | -5 | 350 |
| | Jul | 6830 | 6224 | 53 | 553 |
| | Aug | 6362 | 6227 | -78 | 213 |
| | Sep | 6563 | 6278 | 122 | 164 |
| | Oct | 6581 | 6315 | 251 | 14 |
| | Nov | 6124 | 6335 | 182 | -393 |
| | Dec | 5463 | 6323 | -127 | -733 |
| 2012 | Jan | 5628 | 6282 | -180 | -474 |
| | Feb | 6277 | 6261 | -185 | 201 |
| | Mar | 6522 | 6259 | -176 | 439 |
| | Apr | 6659 | 6281 | 55 | 323 |
| | May | 6500 | 6361 | 88 | 518 |
| | Jun | 6108 | 6518 | -5 | -405 |
| | Jul | 6353 | 6703 | 53 | -402 |
| | Aug | 6337 | 6830 | -78 | -416 |
| | Sep | 6549 | 6895 | 122 | -467 |
| | Oct | 7126 | 6936 | 251 | -61 |
| | Nov | 7504 | 7000 | 182 | 322 |
| | Dec | 7852 | 7071 | -127 | 908 |
| 2013 | Jan | 7663 | 7094 | -180 | 749 |
| | Feb | 7300 | 7075 | -185 | 410 |
| | Mar | 7053 | 7026 | -176 | 203 |
| | Apr | 7115 | 6945 | 55 | 115 |
| | May | 7574 | 6826 | 88 | 660 |
| | Jun | 6737 | 6668 | -5 | 74 |
| | Jul | 6288 | 6497 | 53 | -262 |
| | Aug | 5936 | 6347 | -78 | -333 |
| | Sep | 5788 | 6245 | 122 | -579 |
| | Oct | 5943 | 6190 | 251 | -499 |
| | Nov | 5822 | 6167 | 182 | -527 |
| | Dec | 5750 | 6230 | -127 | –353 |
| 2014 | Jan | 5648 | 6395 | -180 | -568 |
| | Feb | 5718 | 6612 | -185 | -709 |
| | Mar | 6199 | 6891 | -176 | -515 |
| | Apr | 6650 | 7205 | 55 | -609 |
| | May | 7467 | 7523 | 88 | -144 |
| | Jun | 8357 | 7846 | -5 | 516 |
| | Jul | 8647 | 8190 | 53 | 404 |

| | | | | | |
|---|---|---|---|---|---|
| | Aug | 8784 | 8572 | -78 | 290 |
| | Sep | 9616 | 8945 | 122 | 550 |
| | Oct | 9659 | 9287 | 251 | 121 |
| | Nov | 9728 | 9575 | 182 | -30 |
| | Dec | 9615 | 9778 | -127 | -36 |
| 2015 | Jan | 10027 | 9958 | -180 | 249 |
| | Feb | 10502 | 10156 | -185 | 531 |
| | Mar | 10373 | 10294 | -176 | 256 |
| | Apr | 10693 | 10414 | 55 | 224 |
| | May | 10342 | 10597 | 88 | -343 |
| | Jun | 10336 | 10798 | -5 | -457 |
| | Jul | 10985 | | 53 | |
| | Aug | 11208 | | -78 | |
| | Sep | 10498 | | 122 | |
| | Oct | 11671 | | 251 | |
| | Nov | 12097 | | 182 | |
| | Dec | 12075 | | -127 | |

## Analysis of the time series decomposition results

Based on the decomposition work on the time series of the two sectors, we make the following important observations:

From Table 1, we observe that the seasonal components for the Consumer Durables sector index are positive during the period April-May and September- November, with the highest value occurring in the month of November. The seasonal component is the minimum in the month of February every year. The trend values consistently increased over the period 2010 – 2015 albeit with a sluggish rate. The random component has shown considerable fluctuations in its values. However, the trend is the predominant component in the overall time series.

It is natural for the Consumer Durables sector in India to have a dominant seasonal component, as purchase of consumer durable items like air conditioners, refrigerators etc. tend to happen more during the summer period (April-May) and the consumer electronic items like television, micro wave ovens, home theatre systems etc. are sold more during the festive seasons in India which is predominantly during the months of October – November. The companies

in the consumer durables sector in India usually run a number of promotion and price discounts schemes during the festive seasons that lead to increased sales of these items, thus leading to a positive seasonal effect during the festive months.

From Table 2, the time series for the Small Cap sector also is predominantly guided by its trend component. However, when we look at the strength of the random component values with respect to the overall time series value, we observe that in many months, the presence of the random component is quite strong. This also validates our intuition that the Small Cap sector would have a strong random component in its time series.

**Table 2.** *Small Cap Sector Aggregate Index and its Components (2010 – 2015)*

| Year | Month | Aggregate | Trend | Seasonal | Random |
|---|---|---|---|---|---|
| 2010 | Jan | 8677 | | 15 | |
| | Feb | 8230 | | -253 | |
| | Mar | 8448 | | -415 | |
| | Apr | 9053 | | -47 | |
| | May | 8702 | | -123 | |
| | Jun | 8785 | | 43 | |
| | Jul | 9324 | 9324 | 214 | -214 |
| | Aug | 9687 | 9336 | -56 | 408 |
| | Sep | 10158 | 9308 | 135 | 715 |
| | Oct | 10647 | 9275 | 241 | 1131 |
| | Nov | 10544 | 9245 | 279 | 1020 |
| | Dec | 9420 | 9200 | -31 | 251 |
| 2011 | Jan | 9109 | 9134 | 15 | -40 |
| | Feb | 8072 | 9003 | -253 | -677 |
| | Mar | 7942 | 8786 | -415 | -429 |
| | Apr | 8777 | 8502 | -47 | 323 |
| | May | 8256 | 8172 | -123 | 207 |
| | Jun | 8147 | 7850 | 43 | 254 |
| | Jul | 8377 | 7573 | 214 | 590 |
| | Aug | 7482 | 7397 | -56 | 141 |
| | Sep | 7150 | 7295 | 135 | -280 |
| | Oct | 6838 | 7161 | 241 | -564 |

| | | | | | |
|---|---|---|---|---|---|
| | Nov | 6444 | 7002 | 279 | -836 |
| | Dec | 5800 | 6849 | -31 | -1018 |
| 2012 | Jan | 6084 | 6702 | 15 | -634 |
| | Feb | 6871 | 6591 | -253 | 533 |
| | Mar | 6693 | 6532 | -415 | 576 |
| | Apr | 6807 | 6523 | -47 | 331 |
| | May | 6398 | 6563 | -123 | -42 |
| | Jun | 6348 | 6657 | 43 | -351 |
| | Jul | 6649 | 6777 | 214 | -342 |
| | Aug | 6548 | 6823 | -56 | -218 |
| | Sep | 6659 | 6789 | 135 | -264 |
| | Oct | 7122 | 6728 | 241 | 153 |
| | Nov | 7104 | 6681 | 279 | 145 |
| | Dec | 7392 | 6644 | -31 | 779 |
| 2013 | Jan | 7386 | 6579 | 15 | 792 |
| | Feb | 6666 | 6482 | -253 | 437 |
| | Mar | 6081 | 6376 | -415 | 120 |
| | Apr | 5962 | 6266 | -47 | -257 |
| | May | 6101 | 6163 | -123 | 61 |
| | Jun | 5779 | 6070 | 43 | -334 |
| | Jul | 5652 | 5988 | 214 | -550 |
| | Aug | 5227 | 5937 | -56 | -654 |
| | Sep | 5421 | 5951 | 135 | -665 |
| | Oct | 5731 | 6040 | 241 | -550 |
| | Nov | 6009 | 6191 | 279 | -461 |
| | Dec | 6280 | 6448 | -31 | -137 |
| 2014 | Jan | 6504 | 6803 | 15 | -314 |
| | Feb | 6341 | 7191 | -253 | -597 |
| | Mar | 6729 | 7618 | -415 | -474 |
| | Apr | 7454 | 8046 | -47 | -544 |
| | May | 8228 | 8466 | -123 | -116 |
| | Jun | 9815 | 8884 | 43 | 889 |
| | Jul | 10132 | 9283 | 214 | 635 |
| | Aug | 10073 | 9687 | -56 | 442 |
| | Sep | 10819 | 10072 | 135 | 612 |
| | Oct | 10604 | 10416 | 241 | -52 |
| | Nov | 11227 | 10697 | 279 | 251 |
| | Dec | 11072 | 10860 | -31 | 243 |
| 2015 | Jan | 11294 | 10964 | 15 | 315 |
| | Feb | 11255 | 11085 | -253 | 423 |
| | Mar | 11057 | 11146 | -415 | 326 |
| | Apr | 11375 | 11157 | -47 | 266 |

| | | | | |
|---|---|---|---|---|
| May | 11059 | 11172 | -123 | 9 |
| Jun | 10894 | 11191 | 43 | -339 |
| Jul | 11544 | | 214 | |
| Aug | 11578 | | -56 | |
| Sep | 10764 | | 135 | |
| Oct | 10916 | | 241 | |
| Nov | 11294 | | 279 | |
| Dec | 11444 | | -31 | |

In order to investigate further into the behavior of the two time series, we carry out two more experiments. This is driven by our two hypotheses: (i) The Consumer Durables sector displays stronger seasonal characteristics than the Small Cap sector and (ii) The Small Cap sector is dominated by the random component of its time series than the Consumer Durables sector.

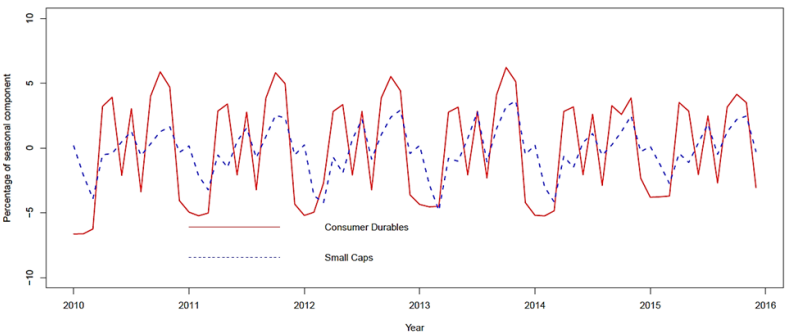


**Figure 5.** *Comparison of the Seasonal Components of the Consumer Durable and the Small Cap Sector (Jan 2010 – Dec 2015)*

Since the absolute values of the time series indices for the two sectors have different scales, it would not make much sense to compare the absolute values of the random and seasonal components of the two time series. Hence, we prepared four text files containing the percentage values of the random and the seasonal components with respect to the overall time series values for both the sectors over the period January 2010 – December 2015. From these four text files, we

created four time series variables in R using the *ts( )* function in the *TTR* package. Using the two seasonal components of the time series (one each for the two sectors), we have created a multiple line plot so that the seasonal components for the two sectors can be visually compared. The same exercise is repeated for the random component time series.

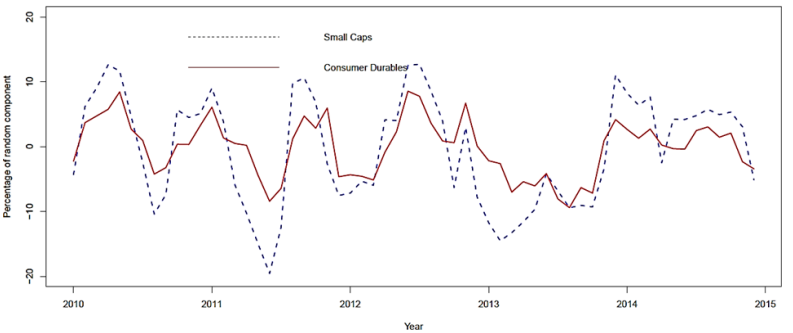


**Figure 6.** *Comparison of the Random Components of the Consumer Durable and the Small Cap Sector (Jan 2010 – Dec 2015)*

Figure 5 depicts the comparison of the percentage of the seasonal components in the overall time series for the Consumer Durables and the Small Cap sector. It can be easily observed that the crests and the troughs of the Consumer Durables sector are much bigger than those of the Small Cap sector. Hence, the results validate our hypothesis (i) – Consumer Durable sector exhibits more seasonality than the Small Cap sector.

Figure 6 presents the comparison of the percentage of the random components in the overall time series for the Consumer Durables and the Small Cap sector. It not difficult to observe that curve for the Small Cap sector has longer amplitude in fluctuations compared to its Consumer Durable counterpart. Hence, the results validate our hypothesis (ii) – Small Cap sector exhibits stronger random

component in its time series than the Consumer Durable sector.

## Proposed frameworks of time series forecasting

In this Section, we discuss some methods that we have applied on the Consumer Durables time series data and the Small Cap time series data for making robust forecasting and for a better understanding of the relative contributions of the constituents (i.e., the trend, seasonal and random components) of a time series. We present five different approaches in forecasting and one method for determining the relative strengths of the trend and seasonality components in a time series.

**Method 1:** The time series data of the Consumer Durables and the Small Cap sectors for the period January 2010 to December 2014 is used for building the forecasting model. The *HoltWinters( )* function in R with changing slope in trend and presence of seasonal components is used to forecast the monthly indices for both the sectors for each month of 2015. The forecasted values are compared with the actual values of the indices and the error of forecast is computed for every month of 2015 for both the sectors. Note that in the approach, the forecast is made in December 2014 for every month of 2015. Therefore, the forecast horizon in the approach is 12 months.

**Method II:** In contrast to Method I, in this approach, the forecasting for each month of 2015 for both the sectors are done on the basis of time series data from January 2010 till the end of the previous month for which the forecast is made. For example, in order to forecast for the month of April 2015, time series data for the period January 2010 to March 2015 are used to build the forecasting model. Hence, this method uses a forecast horizon 1 month. The *HoltWinters( )* function in R with changing slope in trend and presence of seasonal component is used in forecasting. The

errors in forecast are computed for each sector for every month of 2015 in the same manner as in Method I.

**Method III:** The fundamental objective of this approach of forecasting is to investigate how effectively we can forecast the aggregate of the trend and the seasonal components of a times series. Since, the random components in a time series are impossible to predict, we devise an approach of forecasting using the trend and seasonal components of a time series. In this method, we first use the time series data for both the Consumer Durables and the Small Cap sectors for the period January 2010 to December 2014 and decompose both the time series into their respective trend and seasonal and random components. As we have seen in Section 2, the decomposition yields the trend component from July 2010 to June 2014 for each time series, since values for the first six months and last six months are truncated for computations of 12 months' moving averages. Using the computed trends values for both the sectors for the period July 2010 to June 2014, we forecast the trends values for both the sectors for the period January 2015 to June 2015, using *HoltWinters( )* function in R with changing slope in the trend component and a seasonal component (Coghlan, 2015). These forecasted trend values are added to the corresponding monthly seasonal components which were obtained from the decomposition of the time series data for the period January 2010 to December 2014 for both the sectors. These aggregate values of the trend and the seasonal components now constitute our forecasted aggregate trend and seasonal values for both the sectors for the period January 2015 to June 2015. In order to compute the actual aggregates of the trend and seasonal components, we use the time series for both the sectors for the period January 2010 to December 2015 and decompose both the time series into their trend, seasonal and random components. After decomposition, we compute the

aggregate of the actual trend and the actual seasonal values for both the time series for the period January 2015 to June 2015. The errors of forecasts for each month for both the sectors are also computed.

**Method IV:** In this approach, we have used *Auto Regressive Integrated Moving Average* (ARIMA) technique (Coghlan, 2015) for forecasting. Two ARIMA forecast models (one model each for the two sectors – Consumer Durables and Small Cap) are built using the two time series for the Consumer Durables and the Small Cap sectors for the period January 2010 – December 2014. Based on each of these two time series data, we first derive the three parameters of the *Auto Regressive Moving Average* (ARMA) mode, i.e. the *Auto Regression* parameter (p), the *Difference* parameter (d), and the *Moving Average* parameter (q) for both the time series. The values of the three parameters are used to develop the ARIMA models for the two sectors. Finally, the two ARIMA models are used to predict the time series values of the respective sectors for each month of 2015. Since the forecasting for all the months of 2015 is being made in December 2014, the forecast horizon in this ARIMA approach is 12 months.

**Method V:** In this approach, forecasting is done for both the sectors using two ARIMA models (one each for the two sectors) as in Method IV. However, in contrast to Method IV, where we used a forecast horizon of 12 months, in this approach, we have used a forecast horizon of 1 month. Therefore, for forecasting the time series value for each month in 2015, the training data set for building the ARIMA model included time series data from January 2010 till the last month for which the forecast was made. For example, if we need to forecast the time series value for the Consumer Durables sector for the month of June 2015, the training data set for building the ARIMA model would include time series data from January 2010 till May 2015. It is important to note

here that since the training data set for the ARIMA model in this approach is constantly changing (due to inclusion of newer data), it is mandatory to evaluate the ARIMA parameters every time before the forecasting is made for each month of 2015.

**Method VI:** Ideally, in a time series, both the trend and the seasonal components would vary over time. The variation of the random component is also there, However, the variations of the random component is difficult to model and hence the focus of our forecasting approaches is on the variations of the trend and seasonal components. In this approach, we investigate how the seasonal components in the time series vary with time for both the sectors under considerations, e.g., the Consumer Durables sector and the Small Cap sector. For this purpose, we first consider the time series of both the sectors for the period January 2010 to December 2014. Each of these two time series is decomposed into its trend, seasonal and random components and the aggregate values of the trend and the seasonal components for the period July 2010 to June 2014 are computed. It may be noted that the aggregates could not be computed for the periods January 2010 to June 2010 and July 2014 to December 2014 due to truncation of the trend values. Next, we remove the time series values for the period January 2010 to December 2010 from the data under investigation and insert the time series values for the period January 2015 to December 2015. In other words, we now concentrate on the time series values for the period January 2011 to December 2015 for both the sectors. As in the previous case, we again compute the aggregate of the trend and the seasonal components for period July 2010 to June 2014 for both the sectors based on the new time series from January 2011 to December 2015. Since the seasonal component values are expected to change from 2010 to 2015, in order to have an idea about the change in the aggregate values of the trend

and the seasonal components, we compute the percentage of deviation of the computed aggregate of the trend and the seasonal components for both the sector for each month during the period of June 2011 to July 2014, computed based on the two time series (January 2010 – December 2014 and January 2011 – December 2015). In the event of appreciable changes in the seasonal component values, we expect large values of the percentage deviations.

## Forecasting results and analysis

As discussed in Section 4, we have applied five forecasting methods and time series analysis technique on both the Consumer Durables and the Small Cap sectors' time series. In this Section, we present the detailed results and a critical analysis of the relative merits and demerits of each of the forecasting and analysis framework.

**Method I:** As discussed in Section 5, for both the sectors, we make forecast for each month of 2015 based on time series data from January 2010 to December 2014. *HoltWinters( )* function in R library *forecast* has been used with changing slope in trend (i.e., varying trend) and a seasonal component. The *forecast horizon* in the *HoltWinters* model has been chosen to be 12 months for both the sectors so that the forecasted values for all months of 2015 can be obtained. The results of forecasting using Method I for the Consumer Durables sector and the Small Cap sector are presented in Table 3 and Table 4 respectively.

**Table 3.** *Results of Method I of Forecasting for Consumer Durables Sector for the Year 2015*

| Month | Actual Value (A) | Forecasted Value (B) | Error Percentage (B - A) / A *100 |
|---|---|---|---|
| Jan | 10027 | 9451 | 5.74 |
| Feb | 10502 | 9146 | 12.91 |
| Mar | 10373 | 9166 | 11.63 |
| Apr | 10693 | 9647 | 9.78 |
| May | 10342 | 9839 | 4.86 |
| Jun | 10336 | 10076 | 2.52 |
| Jul | 10985 | 9974 | 9.20 |
| Aug | 11208 | 10144 | 9.49 |
| Sep | 10498 | 10649 | 1.44 |
| Oct | 11671 | 10990 | 5.83 |
| Nov | 12097 | 11197 | 7.44 |
| Dec | 12075 | 10859 | 10.07 |

**Table 4.** *Results of Method I of Forecasting for Small Cap Sector for the Year 2015*

| Month | Actual Value (A) | Forecasted Value (B) | Error Percentage (B - A) / A *100 |
|---|---|---|---|
| Jan | 11294 | 10790 | 4.46 |
| Feb | 11255 | 10208 | 9.31 |
| Mar | 11057 | 10105 | 8.61 |
| Apr | 11375 | 10962 | 3.63 |
| May | 11059 | 11396 | 3.05 |
| Jun | 10894 | 11885 | 9.10 |
| Jul | 11544 | 11942 | 3.45 |
| Aug | 11578 | 12000 | 3.64 |
| Sep | 10764 | 12729 | 18.26 |
| Oct | 10916 | 13354 | 22.33 |
| Nov | 11294 | 13904 | 23.11 |
| Dec | 11444 | 13564 | 18.52 |

**Observations:** We observe from Table 3 that the forecasted values closely match the actual values of the Consumer Durables sector even when the forecast horizon is long (12 months). This clearly shows that *HoltWinters* model with changing trend and additive seasonal components is effective in forecasting the Consumer Durable sector time series values for the period 2010 -2015. The error values

exceeded 10% threshold for the months of February, March and December 2015. For the month of February, there was an unexpected increase in the time series value compared to its previous value in January 2015. In March 2015, on the other hand, there was a decrease in the time series index which was also not expected. In December 2015, there was a fall in time series index which was not in tune with the increasing trends in its previous values. The error values in Table 4 indicate that the forecasted values for the Small Caps sector also very closely match its actual values except for the period September – December 2015. The Small Caps time series behaved in a very unpredictable way during this period. A careful look at Table 2 will make it clear that the time series has undergone alternate rise and fall during this period and it was impossible for *HoltWinters* model with a forecast horizon of 12 months to effectively capture this behavior of the time series which resulted in higher values in forecasting errors.

**Method II:** For both the sectors, we have used *HoltWinters( )* function in R with additive seasonal component and a trend with a changing slope. However, in contrast with Method I, the forecast horizon in this method has been chosen to be 1 month. In other words, forecasting is made for each month of 2015 for both the sectors by taking into account time series data from January 2010 till the previous month for which forecasting is being made. Since this approach uses a very small horizon of forecast, it is likely that this method will be able to capture any possible change in trend and seasonal components more effectively than Method I. However, if there is a continuous rise and fall in the time series values, this method may yield worse results compared to those obtained in Method I. The results of forecasting using Method II for the Consumer Durables sector and the Small Cap sector are presented in Table 5 and Table 6 respectively.

**Table 5.** *Results of Method II of Forecasting for Consumer Durables Sector for the Year 2015*

| Month | Actual Value | Forecasted Value | Error Percentage |
|---|---|---|---|
| | (A) | (B) | (B - A) / A *100 |
| Jan | 10027 | 9451 | 5.74 |
| Feb | 10502 | 9655 | 8.07 |
| Mar | 10373 | 10419 | 0.44 |
| Apr | 10693 | 10851 | 1.48 |
| May | 10342 | 10916 | 5.55 |
| Jun | 10336 | 10645 | 2.99 |
| Jul | 10985 | 10274 | 6.47 |
| Aug | 11208 | 11035 | 1.54 |
| Sep | 10498 | 11695 | 11.40 |
| Oct | 11671 | 11001 | 5.74 |
| Nov | 12097 | 11783 | 2.60 |
| Dec | 12075 | 11767 | 2.55 |

**Observations:** We observe from Table 5 that the forecasted values very closely match with the actual values for the Consumer Durable sector. Except for the month of September 2015, the forecast error values have never exceeded the threshold of 10%. The higher error value for the month of September may be attributed to the sudden and unexpected fall in the time series value for that month. The time series was consistently on the rise over the previous few months, and since the forecast horizon is 1 month, *HoltWinters* method expected an increase in the time series value following the increasing trend of the time series. However, the time series value actually decreased and that resulted into a higher value in foresting error. The error values in Table 6 indicate that the forecasted values for the Small Cap sector also very closely match its actual values except for the month of September 2015. The sudden rise in the error value in September 2015 for the Small Caps sector can again be attributed to the sudden decrease in the time series value during that month which was inconsistent with

the increasing trend of the time series in the previous few months.

**Table 6.** *Results of Method II of Forecasting for Small Cap Sector for the Year 2015*

| Month | Actual Value | Forecasted Value | Error Percentage |
|---|---|---|---|
| | (A) | (B) | (B - A) / A *100 |
| Jan | 11294 | 10790 | 4.46 |
| Feb | 11255 | 10669 | 5.21 |
| Mar | 11057 | 11160 | 0.93 |
| Apr | 11375 | 12021 | 5.68 |
| May | 11059 | 11982 | 8.35 |
| Jun | 10894 | 11695 | 7.35 |
| Jul | 11544 | 10960 | 5.06 |
| Aug | 11578 | 11445 | 1.16 |
| Sep | 10764 | 12233 | 13.65 |
| Oct | 10916 | 11504 | 5.39 |
| Nov | 11294 | 11504 | 1.86 |
| Dec | 11444 | 10844 | 5.24 |

**Method III:** In Section 5, we have already discussed the approach followed in this method. We have used the time series data of the Consumer Durables sector from January 2010 to December 2015 to compute the actual values of the trend and the seasonal components. However, since the actual values of trend component are not available for the period July 2015 – December 2015, we concentrate only on the period January 2015 to June 2015 for the purpose of forecasting. The actual trend and seasonal component values and their aggregated monthly values are noted in Columns *B*, *C* and *D* respectively in Table 7. Now, using the time series data for the period January 2010 to December 2014, the trend and the seasonal components are recomputed. Since the trend values during July 2014 to December 2014 will not be available after this computation, we make a forecast for the trend values for the period January 2015 to June 2015 using *HoltWinters* forecasting model with a changing trend

and an additive seasonal component. The forecasted trend values and the past seasonal component values and their corresponding aggregate values are noted in columns *E*, *F* and *G* respectively in Table 7. The error values are also computed and recorded in the rightmost column of the Table 7. For the Small Cap sector time series, the same method has been followed and the results are recorded in Table 8.

**Observation:** The results obtained using Method III for the Consumer Durables and the Small Caps sectors are presented in Table 7 and Table 8 respectively. From Table 7, we observe that the error values have consistently increased from 2.29% in October 2014 to 19.42% in June 2015. Considering the fact that the trend is forecasted over a long period of 12 months (forecasting for July 2014 – June 2015 being done at the end of June 2014) and since the trend component of the Consumer Durables sector had been sluggish over the period of forecast, the forecasting accuracies obtained in Method III can be considered quite satisfactory for the Consumer Durables sector. The fact that the actual trend was not able to keep its pace intact with its forecasted values is evident from the values in the column *B* and the corresponding values in the column *E* of the Table 7.

Since the trend component of the Small Cap sector grew even more sluggishly as compared to the Consumer Durables sector over the period July 2014 – June 2015, the forecast errors for Small Caps sector using Method III have been higher. It is evident from Table 8 that forecast errors has consistent increased from 0.88% in July 2014 to 35.05% in June 2015. The sluggish rate of increase in the actual trend component as compared to the forecasted trend component using the *HoltWinters( )* function with changing slope in trend and an additive seasonal component is evident from the values in the column *B* and the corresponding values in the column *E* of the Table 8. The extreme sluggish rate of growth in the trend component in the Small Cap sector

during the period of July 2014 to June 2015 has rendered Method III not a very effective method of forecasting for the Small Cap sector.

**Table 7.** *Results of Method III of Forecasting for Consumer Durables Sector for the Period July 2014 – June 2015*

| Month | Actual Trend | Actual Seasonal | Actual (Trend + Seasonal) | Forecasted Trend | Past Seasonal | Forecasted (Trend + Seasonal) | % Error |
|---|---|---|---|---|---|---|---|
| | (A) | (B) | (C) | (D) | (E) | (F) | (F-A)/A*100 |
| Jul 2014 | 8190 | 53 | 8243 | 8192 | -12 | 8180 | 0.76 |
| Aug 2014 | 8572 | -78 | 8494 | 8565 | -113 | 8452 | 0.49 |
| Sep 2014 | 8945 | 122 | 9067 | 9030 | 21 | 9051 | 0.18 |
| Oct 2014 | 9287 | 251 | 9538 | 9498 | 258 | 9756 | 2.29 |
| Nov 2014 | 9575 | 182 | 9757 | 9936 | 226 | 10162 | 4.15 |
| Dec 2014 | 9778 | -127 | 9651 | 10327 | -81 | 10246 | 6.17 |
| Jan 2015 | 9958 | -180 | 9778 | 10842 | -206 | 10636 | 8.77 |
| Feb 2015 | 10156 | -185 | 9971 | 11315 | -281 | 11034 | 10.66 |
| Mar 2015 | 10294 | -176 | 10118 | 11713 | -204 | 11509 | 13.75 |
| Apr 2015 | 10414 | 55 | 10469 | 12081 | 35 | 12116 | 15.73 |
| May 2015 | 10597 | 88 | 10685 | 12423 | 210 | 12633 | 18.23 |
| Jun 2015 | 10798 | -5 | 10793 | 12743 | 146 | 12889 | 19.42 |

**Table 8.** *Results of Method III of Forecasting for Small Cap sector for the Period July 2014 – June 2015*

| Month | Actual Trend | Actual Seasonal | Actual (Trend + Seasonal) | Forecastec Trend | Past Seasonal | Forecasted (Trend + Seasonal) | % Error |
|---|---|---|---|---|---|---|---|
| | (A) | (B) | (C) | (D) | (E) | (F) | (F-A)/A*100 |
| Jul 2014 | 9283 | 214 | 9497 | 9293 | 120 | 9413 | 0.88 |
| Aug 2014 | 9687 | -56 | 9631 | 9845 | -102 | 9743 | 1.16 |
| Sep 2014 | 10072 | 135 | 10207 | 10443 | 47 | 10490 | 2.77 |
| Oct 2014 | 10416 | 241 | 10657 | 10996 | 319 | 11315 | 6.17 |
| Nov 2014 | 10697 | 279 | 10976 | 11497 | 281 | 11778 | 7.31 |
| Dec 2014 | 10860 | -31 | 10829 | 11970 | -27 | 11943 | 10.29 |
| Jan 2015 | 10964 | 15 | 10979 | 12840 | 2 | 12842 | 16.97 |
| Feb 2015 | 11085 | -253 | 10832 | 13394 | -294 | 13100 | 20.94 |
| Mar 2015 | 11146 | -415 | 10731 | 13852 | -431 | 13421 | 25.07 |
| Apr 2015 | 11157 | -47 | 11110 | 14247 | -49 | 14198 | 27.79 |
| May 2015 | 11172 | -123 | 11049 | 14605 | -60 | 14545 | 31.64 |
| Jun 2015 | 11191 | 43 | 11234 | 14979 | 193 | 15172 | 35.05 |

**Method IV:** As pointed out in Section 4, we have applied *Auto Regressive Integrated Moving Average* (ARIMA)

technique for the purpose of forecasting on the Consumer Durables and the Small Cap time series data. We have exploited the power of the *auto.arima( )* function defined in the *forecast* package in R for determining the values of the ARIMA parameters for the time series of the two sectors (Coghlan, 2015). For finding the values of the ARIMA parameters for both the time series, we have used the time series for the Consumer Durables sector and the time series for the Small Cap sector for the period January 2010 to December 2014. For the Consumer Durable sector, applying the *auto.arima ( )* function on the time series, we obtain the ARIMA parameters: *Auto Regression* parameter (p) = 0, *Difference* parameter (d) = 1, *Moving Average* parameter (q) = 1. Therefore, the Consumer Durables sector time series for the period January 2010 – December 2014 is designed as an *Auto Regressive Moving Average* (ARMA) model - ARMA(0, 1, 1). From the ARMA(0, 1, 1) model, the corresponding ARIMA model is constructed using the *arima( )* function in R with the two parameters as: (i) Consumer Durables sector time series, (ii) the order of the ARMA for the time series, i.e., (0, 1, 1). Using the resultant ARIMA model, we call the function *forecast. Arima( )* with parameters: (i) the ARIMA model and (ii) the time horizon of forecast. In this method (i.e., Method IV), we make the forecast for each month of the year 2015 based on the time series for the period January 2010 to December 2014, resulting in a forecast horizon of 12 months. The errors in forecasting are also computed. The same approach is also followed for the Small Cap sector. However, the ARMA parameters for the Small Caps time series were found to be (1, 1, 0). The ARIMA model for the Small Caps sector was built according to these values. The results of forecasting using Method IV for the Consumer Durables sector and the Small Cap sector are presented in Table 9 and Table 10 respectively.

**Observation:** From Table 9, it may be observed that the forecast error values are very low for the Consumer Durable sector with the ARIMA method even using a large forecast horizon of 12 months. The three months (May, June and September 2015) for which the error values exceeded the threshold of 10% mark, the time series exhibited unexpected fall as compared to its previous values and hence it was difficult for the ARIMA model with a long forecast horizon of 12 months to predict this behavior. The error values in Table 10 indicate that the Small Cap sector time series for 2015 had a very close fit with the ARIMA model even with a long forecast horizon of 12 months. The maximum error in forecast being less than 5%, the model has provided an excellent framework of forecast for the Small Cap sector. The Small Cap time series for the year 2015 has remained consistent with no abrupt and sudden increase/decrease in its values. This has allowed ARIMA, even with a long forecast horizon of 12 months, to provide a very accurate framework for forecasting.

**Method V:** In this method, we have utilized ARIMA model for forecasting with a forecast horizon of 1 month for both the sectors. The methodology used for constructing the ARIMA models has been the same as it was in Method IV. However, since the model deploys a training data set that is constantly increasing in size due to inclusion of new data, it is mandatory to re-evaluate the ARIMA parameters every time before the *forecast.Arima( )* function is used for the purpose of forecasting. In other words, before forecasting is made for each month of 2015, we compute the values of the ARIMA parameters, and build a new ARIMA model for forecasting. The results of forecasting using Method V for the Consumer Durables sector and the Small Cap sector are presented in Table 11 and Table 12 respectively.

**Observations:** From Table 11, it may be observed that the forecast errors are very low for the Consumer Durables

sector with the ARIMA model using a low forecast horizon of 1 month. This is expected as the small forecast horizon allows the ARIMA model to capture the behavior of the time series more effectively. However, if there is a sharp change in the time series values (i.e. abrupt increase/decrease in the time series values), the ARIMA model with small forecast horizon of 1 month may perform poorly since it assigns highest weight to the last observation. The higher value of forecast error of 12.47% for the month of October 2015 for the Consumer Durables sector may be attributed to this reason. The Consumer Durables sector time series had a fall in its value from August to September which was against the increasing trend of the time series over the last few months. This resulted in a moderate value of 6.77% of the forecast error for the month of September 2015. If this fall continued in the month of October 2015, ARIMA would have provided a very low value of the forecast error. However, in the month of October 2015, the time series exhibited an increase in its value resulting in a high value of the error rate. Since the time series of the Small Cap sector did not exhibit any abrupt increase or fall in its values in the year 2015, ARIMA model with a forecast horizon of 1 month has produced consistently low error rates in forecasting as evident from Table 12.

**Table 9.** *Results of Method IV of Forecasting for Consumer Durables Sector for the Year 2015*

| Month | Actual Value | Forecasted Value | Error Percentage |
|---|---|---|---|
| | (A) | (B) | (B -A)/A *100 |
| Jan | 10027 | 10232 | 2.04 |
| Feb | 10502 | 10743 | 2.29 |
| Mar | 10373 | 11095 | 6.96 |
| Apr | 10693 | 11381 | 6.43 |
| May | 10342 | 11628 | 12.43 |
| Jun | 10336 | 11849 | 14.63 |
| Jul | 10985 | 12050 | 9.70 |
| Aug | 11208 | 12237 | 9.18 |

| | | | |
|---|---|---|---|
| Sep | 10498 | 12411 | 18.22 |
| Oct | 11671 | 12576 | 7.75 |
| Nov | 12097 | 12732 | 5.24 |
| Dec | 12075 | 12881 | 6.67 |

**Table 10.** *Results of Method IV of Forecasting for Small Cap Sector for the Year 2015*

| Month | Actual Value | Forecasted Value | Error Percentage |
|---|---|---|---|
| | (A) | (B) | (B-A) / A *100 |
| Jan | 11294 | 11030 | 2.34 |
| Feb | 11255 | 11018 | 2.11 |
| Mar | 11057 | 11015 | 0.38 |
| Apr | 11375 | 11014 | 3.17 |
| May | 11059 | 11014 | 0.41 |
| Jun | 10894 | 11014 | 1.10 |
| Jul | 11544 | 11014 | 4.59 |
| Aug | 11578 | 11014 | 4.87 |
| Sep | 10764 | 11014 | 2.32 |
| Oct | 10916 | 11014 | 0.90 |
| Nov | 11294 | 11014 | 2.48 |
| Dec | 11444 | 11014 | 3.76 |

**Summary of Performance:** In Table 13 and Table 14, we have summarized the performance of the five forecasting approaches that we have discussed so far for the Consumer Durables sector and the Small Cap sector respectively. For the purpose of comparison, we have chosen four metrics: (i) minimum (min) error rate, (ii) maximum (max) error rate, (iii) mean error rate, and (iv) standard deviation (sd) of error rates.

As observed from Table 13, for the Consumer Durables sector, Method V that used ARIMA with a forecast horizon of 1 month, has performed best in two metrics: (i) min error rate and (ii) mean error rate. Method II that used *HoltWinters* forecasting model with a forecast horizon of 1 month has performed best in the other two metrics: (i) max error rate and (ii) sd of error rates. Method I that used *HoltWinters* forecasting models with a forecast horizon of 12 months has

performed next with a mean error rate of 7.58 and an sd of error rate of 3.56. Method III that used aggregate of trend and seasonal components has been ranked at fourth position based on the mean error value of 8.38 and Method IV that used ARIMA with a forecast horizon of 12 months has performed worst with a mean error value of 8.46. It is evident, that for the Consumer Durables time series in the year 2015, forecasting methods that used small forecast horizons produced better results. This is due to the fact that the time series did not exhibit abrupt increase/decrease in its values over successive months over the period of forecast.

From Table 14, for the Small Cap sector, it is evident that the Method IV that used ARIMA with a forecast horizon of 12 months have performed best in all the four metrics: (i) min, (ii) max, (iii) mean, and (iv) sd of error rates. Method V that used ARIMA with a forecast horizon of 1 month has been the next in terms of performance with respect to all the four metrics. Since the time series had exhibited changes in its behavior in quite a number of instances in 2015, ARIMA with forecast horizon of 1 month could not provide the best results. However, since the magnitude of those changes were very nominal, the ARIMA with a long horizon of 12 months produced best forecasting results, and the ARIMA with a forecast horizon of 1 month providing good results too. Method II using *HoltWinters( )* function with a forecast horizon of 1 month and Method I using *HoltWinters( )* function with a forecast horizon of 12 months were next in terms of their mean percentage error values. As expected, Method III that used aggregate of the forecasted trend and seasonal values, produced the worst forecasting results for the Small Cap sector. The trend component of the Small Cap sector has been very slow in its growth resulting in a large gap between actual trend values and the forecasted trend values for the year 2015. This gap is responsible for the high

mean percentage error (15.50) in forecasting in Method III for the Small Cap sector.

**Table 11.** *Results of Method V of Forecasting for Consumer Durables Sector for the Year 2015*

| Month | Actual Value | Forecasted Value | Error Percentage |
|---|---|---|---|
| | (A) | (B) | (B - A) / A *100 |
| Jan | 10027 | 10231 | 2.03 |
| Feb | 10502 | 10218 | 2.70 |
| Mar | 10373 | 10615 | 2.33 |
| Apr | 10693 | 10272 | 3.94 |
| May | 10342 | 10861 | 5.02 |
| Jun | 10336 | 10214 | 1.18 |
| Jul | 10985 | 10334 | 5.93 |
| Aug | 11208 | 11206 | 0.02 |
| Sep | 10498 | 11209 | 6.77 |
| Oct | 11671 | 10216 | 12.47 |
| Nov | 12097 | 12043 | 0.45 |
| Dec | 12075 | 12108 | 0.27 |

**Table 12.** *Results of Method V of Forecasting for Small Cap Sector for the Year 2015*

| Month | Actual Value | Forecasted Value | Error Percentage |
|---|---|---|---|
| | (A) | (B) | (B - A) / A *100 |
| Jan | 11294 | 11030 | 2.34 |
| Feb | 11255 | 11844 | 5.23 |
| Mar | 11057 | 11245 | 1.70 |
| Apr | 11375 | 11004 | 3.26 |
| May | 11059 | 11459 | 3.62 |
| Jun | 10894 | 10978 | 0.77 |
| Jul | 11544 | 10852 | 5.99 |
| Aug | 11578 | 11706 | 1.11 |
| Sep | 10764 | 11586 | 7.64 |
| Oct | 10916 | 10668 | 2.27 |
| Nov | 11294 | 11027 | 2.36 |
| Dec | 11444 | 11296 | 1.29 |

**Table 13.** *Performance of the Forecasting Methods for the Consumer Durables Sector*

| Metrics / Methods | Min Error | Max Error | Mean Error | SD of Errors |
|---|---|---|---|---|
| Method 1 | 1.44 | 12.91 | 7.58 | 3.56 |
| Method II | 0.44 | 11.40 | 4.55 | 3.20 |
| Method III | 0.18 | 19.42 | 8.38 | 7.10 |
| Method IV | 2.04 | 18.22 | 8.46 | 4.78 |
| Method V | 0.02 | 12.47 | 3.59 | 3.58 |

**Table 14.** *Performance of the Forecasting Methods for the Small Cap Sector*

| Metrics / Methods | Min Error | Max Error | Mean Error | SD of Errors |
|---|---|---|---|---|
| Method 1 | 3.05 | 23.11 | 10.62 | 7.78 |
| Method II | 0.93 | 13.65 | 5.36 | 3.47 |
| Method III | 0.88 | 35.05 | 15.50 | 12.36 |
| Method IV | 0.38 | 4.87 | 2.37 | 1.52 |
| Method V | 0.77 | 7.64 | 3.13 | 2.14 |

**Table 15.** *Computation Results Using Method VI - Structural Analysis of Trend and Seasonal Components of the Consumer Durables Sector Index for the Period: July 2011 – June 2014*

| Year | Month | Computation 1 (Based on 2010-2014) | | | Computation 2 (Based on 2011-2015) | | | % Variation |
|---|---|---|---|---|---|---|---|---|
| | | Trend | Seasonal | Sum | Trend | Seasonal | Sum | |
| | | A | B | C = A + B | D | E | F = D + E | (F - C)/C *100 |
| 2011 | Jul | 6224 | -12 | 6212 | 6224 | 142 | 6366 | 2.48 |
| | Aug | 6227 | -113 | 6114 | 6227 | -123 | 6104 | 0.16 |
| | Sep | 6278 | 21 | 6299 | 6278 | 54 | 6332 | 0.52 |
| | Oct | 6315 | 258 | 6573 | 6315 | 161 | 6476 | 1.48 |
| | Nov | 6335 | 226 | 6561 | 6335 | 42 | 6377 | 2.80 |
| | Dec | 6323 | -81 | 6242 | 6323 | -164 | 6159 | 1.33 |
| 2012 | Jan | 6282 | -206 | 6076 | 6282 | 175 | 6107 | 0.51 |
| | Feb | 6261 | -281 | 5980 | 6261 | -61 | 6200 | 3.67 |
| | Mar | 6259 | -204 | 6055 | 6259 | -65 | 6194 | 2.30 |
| | Apr | 6281 | 35 | 6316 | 6281 | 84 | 6365 | 0.78 |
| | May | 6361 | 210 | 6571 | 6361 | 160 | 6521 | 0.76 |
| | Jun | 6518 | 146 | 6664 | 6518 | -57 | 6461 | 3.05 |
| | Jul | 6703 | -12 | 6691 | 6703 | 142 | 6845 | 2.31 |
| | Aug | 6830 | -113 | 6717 | 6830 | -123 | 6707 | 0.15 |
| | Sep | 6895 | 21 | 6916 | 6895 | 54 | 6949 | 0.48 |
| | Oct | 6936 | 258 | 7194 | 6936 | 161 | 7097 | 1.35 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Nov | 7000 | 226 | 7226 | 7000 | 42 | 7042 | 2.55 |
| | Dec | 7071 | -81 | 6990 | 7071 | -164 | 6907 | 1.19 |
| 2013 | Jan | 7094 | -206 | 6888 | 7094 | 175 | 6919 | 0.45 |
| | Feb | 7075 | -281 | 6794 | 7075 | -61 | 7014 | 3.24 |
| | Mar | 7026 | -204 | 6822 | 7026 | -65 | 6961 | 2.04 |
| | Apr | 6945 | 35 | 6980 | 6945 | 84 | 7029 | 0.70 |
| | May | 6826 | 210 | 7036 | 6826 | 160 | 6986 | 0.71 |
| | Jun | 6668 | 146 | 6814 | 6668 | -57 | 6611 | 2.98 |
| | Jul | 6497 | -12 | 6485 | 6497 | 142 | 6639 | 2.38 |
| | Aug | 6347 | -113 | 6234 | 6347 | -123 | 6224 | 0.16 |
| | Sep | 6245 | 21 | 6266 | 6245 | 54 | 6299 | 0.53 |
| | Oct | 6190 | 258 | 6448 | 6190 | 161 | 6351 | 1.50 |
| | Nov | 6167 | 226 | 6393 | 6167 | 42 | 6209 | 2.89 |
| | Dec | 6230 | -81 | 6149 | 6230 | -164 | 6066 | 1.35 |
| 2014 | Jan | 6395 | -206 | 6189 | 6395 | -175 | 6220 | 0.50 |
| | Feb | 6612 | -281 | 6331 | 6612 | -61 | 6551 | 3.47 |
| | Mar | 6891 | -204 | 6687 | 6891 | -65 | 6826 | 2.08 |
| | Apr | 7205 | 35 | 7240 | 7205 | 84 | 7289 | 0.68 |
| | May | 7523 | 210 | 7733 | 7523 | 160 | 7683 | 0.65 |
| | Jun | 7846 | 146 | 7992 | 7846 | -57 | 7789 | 2.54 |

**Table 16.** *Computation Results using Method VI - Structural Analysis of Trend and Seasonal Components of the Small Cap Sector Index for the Period: July 2011 – June 2014*

| Year | Month | Computation 1 (Based on 2010-2014) | | | Computation 2 (Based on 2011-2015) | | | % Change |
|---|---|---|---|---|---|---|---|---|
| | | Trend | Seasonal | Sum | Trend | Seasonal | Sum | |
| | | A | B | C = A + B | D | E | F = D + E | (F - C)/C *100 |
| 2011 | Jul | 7573 | 120 | 7693 | 7573 | 329 | 7902 | 2.72 |
| | Aug | 7397 | -102 | 7295 | 7397 | -97 | 7300 | 0.07 |
| | Sep | 7295 | 47 | 7342 | 7295 | 17 | 7312 | -0.41 |
| | Oct | 7161 | 319 | 7480 | 7161 | 19 | 7180 | -4.01 |
| | Nov | 7002 | 281 | 7283 | 7002 | 85 | 7087 | -2.69 |
| | Dec | 6849 | -27 | 6822 | 6849 | -33 | 6816 | -0.09 |
| 2012 | Jan | 6702 | 2 | 6704 | 6702 | 87 | 6789 | 1.27 |
| | Feb | 6591 | -294 | 6297 | 6591 | -23 | 6568 | 4.30 |
| | Mar | 6532 | -431 | 6101 | 6532 | -246 | 6286 | 3.03 |
| | Apr | 6523 | -49 | 6474 | 6523 | -67 | 6456 | -0.28 |
| | May | 6563 | -60 | 6503 | 6563 | -113 | 6450 | -0.82 |
| | Jun | 6657 | 193 | 6850 | 6657 | 40 | 6697 | -2.23 |
| | Jul | 6777 | 120 | 6897 | 6777 | 329 | 7106 | 3.03 |
| | Aug | 6823 | -102 | 6721 | 6823 | -97 | 6726 | 0.07 |
| | Sep | 6789 | 47 | 6836 | 6789 | 17 | 6806 | -0.44 |
| | Oct | 6728 | 319 | 7047 | 6728 | 19 | 6747 | -4.26 |
| | Nov | 6681 | 281 | 6962 | 6681 | 85 | 6766 | -2.82 |
| | Dec | 6644 | -27 | 6617 | 6644 | -33 | 6611 | -0.09 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 2013 | Jan | 6579 | 2 | 6581 | 6579 | 87 | 6666 | 1.29 |
| | Feb | 6483 | -294 | 6189 | 6483 | -23 | 6460 | 4.38 |
| | Mar | 6376 | -431 | 5945 | 6376 | -246 | 6130 | 3.11 |
| | Apr | 6266 | -49 | 6217 | 6266 | -67 | 6199 | -0.29 |
| | May | 6163 | -60 | 6103 | 6163 | -113 | 6050 | -0.87 |
| | Jun | 6071 | 193 | 6264 | 6071 | 40 | 6111 | -2.44 |
| | Jul | 5988 | 120 | 6108 | 5988 | 329 | 6317 | 3.42 |
| | Aug | 5938 | -102 | 5836 | 5938 | -97 | 5841 | 0.09 |
| | Sep | 5951 | 47 | 5998 | 5951 | 17 | 5968 | -0.50 |
| | Oct | 6040 | 319 | 6359 | 6040 | 19 | 6059 | -4.72 |
| | Nov | 6191 | 281 | 6472 | 6191 | 85 | 6276 | -3.03 |
| | Dec | 6448 | -27 | 6421 | 6448 | -33 | 6415 | -0.09 |
| 2014 | Jan | 6803 | 2 | 6805 | 6803 | 87 | 6890 | 1.25 |
| | Feb | 7191 | -294 | 6897 | 7191 | -23 | 7168 | 3.92 |
| | Mar | 7618 | -431 | 7187 | 7618 | -246 | 7372 | 2.57 |
| | Apr | 8046 | -49 | 7997 | 8046 | -67 | 7979 | -0.23 |
| | May | 8466 | -60 | 8406 | 8466 | -113 | 8353 | -0.63 |
| | Jun | 8884 | 193 | 9077 | 8884 | 40 | 8924 | -1.69 |

**Method VI:** The objective of this method is to gain an insight into the contribution of the trend and the seasonal components in the overall time series of the Consumer Durables and the Small Cap sector. As we mentioned in this Section 5, this approach is based on comparison of the aggregate of the trend and the seasonal components of a time series over two different period of time. First, we construct a time series using the time series data for the period January 2010 to December 2014, and then we compute the trend and the seasonal components and their aggregate values. We refer to this computation as *Computation 1*. For the Consumer Durable sector time series, the trend, the seasonal and their aggregate values in *Computation 1* are noted in the columns *A*, *B* and *C* respectively in Table 15. Next, we construct the second time series using the time series data for the period January 2011 to December 2015 and repeat the computation of the trend, the seasonal and their aggregate values. We refer to this computation as *Computation 2*. For the Consumer Durables sector, the trend, the seasonal and their aggregate values in *Computation 2* are noted in the columns *D*, *E* and *F*

respectively in Table 15. The percentages of variation of the aggregate values in both computations are noted for each month for the period July 2011 to June 2014. If there is a structural difference between the time series data in 2010 and 2015, then we expect that difference to be reflected in the aggregate of the trend and the seasonal values. The computations for the Small Cap sector are presented in Table 16.

**Observation:** From both Table 15 and Table 16, it is quite evident that the aggregate of the trend and the seasonal components had remained consistently the same over the period July 2011 to June 2014 for both the Consumer Durables and the Small Cap sector. This indicates that there have been no structural changes in the time series of these two sectors during the period January 2010 to December 2015. Since the change of the time series due to substitution of the 2010 data by 2015 data had virtually no impact on the trend and the seasonal components, we conclude that the impact of the random component is not significant, and the Consumer Durables and the Small Cap sectors time series is quite amenable for robust forecasting.

## Related Work

Researchers have spent considerable effort in designing mechanisms for forecasting of daily stock prices. Applications of neural network based approaches have been proposed in many forecasting systems. Mostafa proposed neural network-based mechanism to predict stock market movements in Kuwait (Mostafa, 2010). Kimoto et al applied neural networks on historical accounting data and used various macroeconomic parameters for the purpose of prediction of variations in stock returns (Kimoto *et al*, 1990). Leigh et al proposed the use of linear regression and simple neural network models for forecasting the stock market indices in the New York Stock Exchange during the period

1981-1999 (Leigh *et al,* 2005). Hammad et al have demonstrated that artificial neural network (ANN) model can be trained to converge to an optimal solution while it maintains a very high level of precision in forecasting of stock prices (Hammad *et al,* 2009). Dutta et al demonstrate the application of ANN models for forecasting Bombay Stock Exchange's SENSEX weekly closing values for the period of January 2002-December 2003 (Dutta *et al,* 2006). Tsai and Wang found observations that highlighted the fact that Bayesian Network-based approaches have better forecasting power than traditional regression and neural network-based approaches (Tsai & Wang, 2009). Tseng et al deployed traditional time series decomposition (TSD), HoltWinters (H/W) models, Box-Jenkins (B/J) methodology and neural network- based approach on 50 randomly chosen stocks during September 1, 1998 - December 31, 2010 for forecasting the future values of the stock prices (Tseng *et al,* 2012). It has been observed that forecasting errors are lower for B/J, H/W and normalized neural network model, while the errors are appreciably larger for time series decomposition and non-normalized neural network models. Moshiri and Cameron presented a Back Propagation Network (BPN) with econometric models to forecast inflation using (i) Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) model, (ii) Vector Autoregressive (VAR) model and (iii) Bayesian Vector Autoregressive (BVAR) model (Moshiri, & Cameron, 2010). Senol & Ozturan illustrated that ANN can be used to predict stock prices and their direction of changes (Senol & Ozturan, 2008). The result was promising with a forecast accuracy of 81% on the average. Hutichinson et al proposed a non-parametric method for estimating the pricing formula of a derivative that applied the principles of *learning networks* (Hutchinson *et al,* 1994). The inputs to the network were the primary economic variables that influence the derivative price, e.g.,

the current fundamental asset price, the strike price, the time to maturity etc. The derivative price was defined to be the output into which the learning network maps the inputs. The data used were the daily closing prices of S&P 500 futures and the options for the 5-year period from January 1987 to December 1991. The authors have compared their results with the parametric derivative pricing formula and the found the results quite promising. Thenmozhi examined the nonlinear nature of the Bombay Stock Exchange time series using chaos theory (Thenmozhi, 2001). The study examined the Sensex returns time series from August 1980 to September 1997 and showed that the daily returns and weekly returns of the BSE sensex are characterized by nonlinearity and the time series is weakly chaotic.

ANN and Hybrid systems are particularly effective in forecasting stock prices for stock time series data. A large number of work have been done based on ANN techniques for stock market prediction (Shen *et al,* 2007; Jaruszewicz & Mandziuk, 2004; Ning *et al,* 2009; Pan *et al,* 2005; Hamid & Iqbal, 2004; Chen, *et al,* 2005; Chen *et al,* 2003; Hanias *et al,* 2012; de Faria, 2009). Many applications of hybrid systems in stock market time series data analysis have also been proposed in the literature (Wu et al, 2008; Wang & Nie, 2008; Perez-Rodriguez *et al,* 2005; Leung *et al,* 2000; Kim, 2004).

In contrast to the work mentioned above, our approach in this chapter is based on structural decomposition of a time series to study the behavior of two different sectors of the Indian economy – the Consumer Durables sector and the Small Cap sector. By decomposition of the time series of these two sectors for the period January 2010 – December 2015, we have demonstrated the fundamental differences between them. We found that while the seasonal component is much stronger in the Consumer Durables sector time series, the time series of the Small Cap sector had a dominant random component. Besides illustrating the fundamental

differences between these time series, we have proposed five robust forecasting techniques and a quantitative framework for analyzing any change in behavior of the constituents of a time series over a long period of time so as to have an idea how effectively the time series future values may be predicted. We have computed the relative accuracies of each of the forecasting techniques, and also have critically analyzed under what situations a particular technique performs better than the other techniques. Our proposed framework of analysis can be used as a broad approach for forecasting the behavior of other stock market indices in India.

## Conclusion

In this chapter, we proposed a time series decomposition-based approach for deeper understanding and analysis of two sectors of the Indian economy – the Consumer Durables sector and the Small Cap sector. We have demonstrated that decomposition results provide us insights about the fundamental characteristics of the sectors which in turn can enable the investors in making wise and efficient investment decisions about their portfolios. Using our proposed decomposition approach, we have also validated two hypotheses – (i) the Consumer durables sector has a strong seasonal component and (ii) the Small Cap sector is characterized by the presence of a strong random component. After analyzing the time series decomposition results and validating the hypotheses, we have proposed five robust forecasting techniques and a quantitative framework for analyzing the behavior of the structural constituents of a time series. We have presented detailed results on the performance of each of the forecasting methods and also critically analyzed why certain method has performed best compared to the others, for what type of time series and under what situations. The proposed

structural decomposition and analysis approach provided enough insights about the way the constituents of the time series for the two sectors had behaved over the period under investigation, i.e., January 2010 – December 2015. It has been demonstrated clearly that the time series of both the sectors are quite amenable for robust and accurate forecasting even in presence of a dominant random component.

The results obtained from the above analysis are extremely useful for portfolio construction. When we perform this analysis for other sectors as well, it will help portfolio managers and individual investors to identify which sector, and in turn which stock, to buy/sell in which period. It will also help in identifying which sector, and hence which stock, is dominated by the random component and thus is speculative in nature.

## References


Chen, A.-S., Leung, M.T., & Daouk, H. (2003). Application of neural networks to an emerging financial market: Forecasting and trading the Taiwan Stock Index. *Operations Research in Emerging Economics*, 30(6), 901-923. doi. 10.1016/S0305-0548(02)00037-0

Chen, Y., Dong, X., & Zhao, Y. (2005). Stock index modeling using EDA based local linear wavelet neural network. *Proceedings of International Conference on Neural Networks and Brain*, 1646-1650. doi. 10.1109/ICNNB.2005.1614946

Coghlan, A. (2015). *A Little Book of R for Time Series*, Release 02. [Retrieved from].

Data-Chaudhuri, T., Ghosh, I., & Eram, S. (2016). Predicting stock returns of mid cap firms in India: an application of random forest and dynamic evolving neural fuzzy inference system. *Proceedings of the 2nd National Conference on Advances in Business Research and Management Practices (ABRMP'16)*, Kolkata, India, January 8-9, 2016. doi. 10.2139/ssrn.2709913.

de Faria, E., Albuquerque, M.P., Gonzalez, J., Cavalcante, J. & Albuquerque, M.P. (2009). Predicting the Brazilian stock market through neural networks and adaptive exponential smoothing methods. *Expert Systems with Applications*, 36(10), 12506-12509. doi. 10.1016/j.eswa.2009.04.032

Dutta, G., Jha, P., Laha, A. & Mohan, N. (2006). Artificial neural network models for forecasting stock price index in the Bombay stock exchange. *Journal of Emerging Market Finance*, 5, 283-295. doi. 10.1177/097265270600500305

Hamid, S.A., & Iqbal, Z. (2004). Using neural networks for forecasting volatility of S&P 500 index futures prices. *Journal of Business Research*, 57(10), 1116–1125. doi. 10.1016/S0148-2963(03)00043-2

Hammad, A., Ali, S. & Hall, E. (2009). Forecasting the Jordanian Stock Price Using Artificial Neural Network. [Retrieved from].

Hanias, M., Curtis, P. & Thalassinos, J. (2012). Time Series Prediction with Neural Networks for the Athens Stock Exchange Indicator. *European Research Studies*, 15(2), 23-31.

Hutchinson, J. M., Lo, A. W., & Poggio, T. (1994). A nonparametric approach to pricing and hedging derivative securities via learning networks. *Journal of Finance*, 49(3), 851-889. doi. 10.3386/w4718

Ihaka, R. & Gentleman, R. (1996). A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299-314. doi. 10.2307/1390807

Jaruszewicz, M. & Mandziuk, J. (2004). One day prediction of NIKKEI index considering information from other stock markets. *Proceedings of*

*the International Conference on Artificial Intelligence and Soft Computing*, 3070, 1130-1135. doi. 10.1007/978-3-540-24844-6_177

Kimoto, T., Asakawa, K., Yoda, M. & Takeoka, M. (1990). Stock market prediction system with modular neural networks. *Proceedings of the IEEE International Joint Conference on Neural Networks(IJCNN)*, 1-16. doi. 10.1109/IJCNN.1990.137535

Leigh, W., Hightower, R. & Modani, N. (2005). Forecasting the New York Stock Exchange composite index with past price and interest rate on condition of volume spike. *Expert Systems with Applications*, 28(1), 1-8. doi. 10.1016/j.eswa.2004.08.001

Leung, M.T., Daouk, H. & Chen, A.-S. (2000). Forecasting stock indices: A comparison of classification and level estimation models. *International Journal of Forecasting*, 16(2), 173-190. doi. 10.2139/ssrn.200429

Moshiri, S. & Cameron, N. (2010). Neural network versus econometric models in forecasting inflation. *Journal of Forecasting*, 19(3), 201-217. doi. 10.1002/(SICI)1099-131X(200004)19:3<201::AID-FOR753>3.0.CO;2-4

Mostafa, M. (2010). Forecasting stock exchange movements using neural networks: Empirical evidence from Kuwait. *Expert Systems with Application*, 37(9), 6302-6309. doi. 10.1016/j.eswa.2010.02.091

Ning, B., Wu, J., Peng, H. & Zhao, J. (2009). Using chaotic neural network to forecast stock index. *Advances in Neural* Networks, 5551, 870-876. doi. 10.1007/978-3-642-01507-6_98

Pan, H., Tilakaratne, C. & Yearwood, J. (2005). Predicting the Australian stock market index using neural networks exploiting dynamical swings and intermarket influences. *Advances in Artificial Intelligence*, Lecture Notes in Computer Science, Springer-Verlag, 2903, 327-338 doi. 10.1007/978-3-540-24581-0_28

Perez-Rodriguez, J.V., Torra, S. & Andrada-Felix, J. (2005). Star and ANN models: Forecasting performance on the Spanish IBEX-35 stock index. *Journal of Empirical Finance*, 12(3), 490–509. doi. 10.1016/j.jempfin.2004.03.001

Sen, J. & Datta-Chaudhuri, T. (2016a). Decomposition of time series data of stock markets and its implications for prediction - An application for the Indian auto sector. *Proceedings of the 2nd National Conference on Advances in Business Research and Management Practices (ABRMP'16)*, Kolkata, India, January 8 -9, 2016. [Retrieved from].

Sen, J. & Datta-Chaudhuri, T. (2016b). A framework for predictive analysis of stock market indices - A study of the Indian auto sector. *Calcutta Business School (CBS) Journal of Management Practices*, 2(2), 1-19. [Retrieved from].

Senol, D. & Ozturan, M. (2008). Stock price direction prediction using artificial neural network approach: The case of Turkey. *Journal of Artificial Intelligence*, 1, 70-77. doi. 10.3923/jai.2008.70.77

Shen, J., Fan, H. & Chang, S. (2007). Stock index prediction based on adaptive training and pruning algorithm. *Advances in Neural Networks*, Lecture Notes in Computer Science, Springer-Verlag, 4492, 457–464. doi. 10.1007/978-3-540-72393-6_55

Thenmozhi, M. (2006). Forecasting stock index numbers using neural networks. *Delhi Business Review*, 7(2), 59-69. [Retrieve from].

Tsai, C.-F. & Wang, S.-P. (2009). Stock price forecasting by hybrid machine learning techniques. *Proceedings of International Multi Conference of Engineers and Computer Scientists*, 1. [Retrieved from].

Tseng, K-C., Kwon, O., & Tjung, L.C. (2012). Time series and neural network forecast of daily stock prices. *Investment Management and Financial Innovations*, 9(1), 32-54. [Retrieved from].

Wang, W. & Nie, S. (2008). The performance of several combining forecasts for stock index. *International Workshop on Future Information Technology and Management Engineering*, 450-455. doi. 10.1109/FITME.2008.42

Wu, Q., Chen, Y. & Liu, Z. (2008). Ensemble model of intelligent paradigms for stock market forecasting. *Proceedings of the IEEE 1st International Workshop on Knowledge Discovery and Data Mining*, 205 – 208, Washington, DC, USA. doi. 10.1109/WKDD.2008.54

Zhu, X., Wang, H., Xu, L. & Li, H. (2008). Predicting stock index increments by neural networks: The role of trading volume under different horizons. *Expert Systems Applications*, 34(4), 3043-3054. doi. 10.1016/j.eswa.2007.06.023

**For Cited this Chapter:**

**Sen, J., & Chaudhuri, T.D. (2020). An alternative framework for time series decomposition and forecasting and its relevance for portfolio choice - A comparative study of the Indian consumer durable and small cap sectors, *In* T.D. Chaudhuri (Ed.), *Overview of the Indian Economy: Empirical analysis of financial markets, exchange rate movements and industrial development* (pp.45-87), KSP Books: Istanbul.**

ISBN: 978-605-7736-93-2 (e-Book)
KSP Books 2020
© KSP Books 2020

**Copyrights**

Copyright for this Book is retained by the author(s), with first publication rights granted to the Book. This is an open-access Book distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by-nc/4.0 ).

# 4 Using clustering method to understand Indian stock market volatility

*By*
Tamal DATTA CHAUDHURI
& Indranil GHOSH

## Introduction

The reasons for studying stock market volatility are that it i) aids in intraday trading, ii) is the basis of neutral trading in the options market, iii) affects portfolio rebalancing by fund managers, iv) helps in hedging, v) affects capital budgeting decisions through timing of raising equity from the market and its pricing and also vi) affects policy decisions relating to the financial markets. Extensive research has been done on stock market volatility and its implications, the thrust being on forecasting volatility. The measures that have been used for estimating volatility are historic volatility and implied volatility.

The literature has used econometric techniques like ARCH, GARCH models to estimate volatility. Using the mean reversal property of volatility, researchers have used decile analysis to predict volatility. This is useful for options traders. There has been application of Artificial Neural

Network (ANN) models to forecast stock market volatility. This chapter explores the role of Clustering Algorithms in forecasting volatility. We go beyond simply forecasting volatility and ask the question as to whether stock market volatility can be predicted at all and if so, within what time bounds. That is, whether it is meaningful to take a long time series data and predict volatility, without understanding the pattern in the data and its characteristics.

If we focus on implied volatility and study the data on India VIX, the implied volatility index in India, for the period 2008 to 2015 (June), Figure 1 shows that there is no specific trend or pattern in this data for long term forecasting. There are spikes in the data, and if we club the entire data for our analysis, we may be erring. Instead we suggest Clustering Algorithms in this chapter to identify patterns in the data. For our analysis we map number of clusters against number of variables. We then test for efficiency of clustering. Our contention is that, given a fixed number of variables, one of them being historic volatility of NIFTY returns, if increase in the number of clusters improves clustering efficiency, then volatility cannot be predicted. Volatility then becomes random as, for a given time period, it gets classified in various clusters. On the other hand, if efficiency falls with increase in the number of clusters, then volatility can be predicted as there is some homogeneity in the data. Further, if we fix the number of clusters and then increase the number of variables, this should have some impact on clustering efficiency. Indeed if we can hit upon, in a sense, an optimum number of variables, then if the number of clusters is reasonably small, we can use these variables to predict volatility.

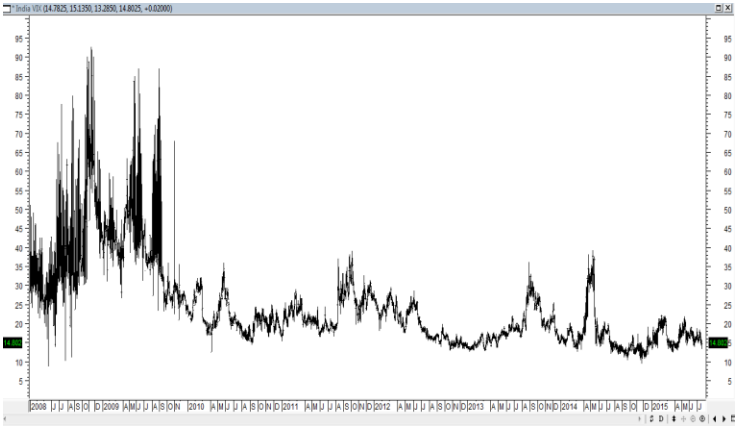

**Figure 1.** *India VIX for the period of 2008-2015 (June)*
**Source:** Metastock

## Objective of the chapter

The objective of this chapter is to present a framework of analysis based on Clustering Algorithms to forecast stock market volatility. The variables that we consider for our study are volatility of NIFTY returns, volatility of gold returns, India VIX, CBOE VIX, volatility of crude oil returns, volatility of DJIA returns, volatility of DAX returns, volatility of Hang Seng returns and volatility of Nikkei returns. Three clustering algorithms namely Kernel K-Means, Self-Organizing Maps and Mixture of Gaussian models will be used to carry out the clustering operation and two internal clustering validity measures, Silhouette Index and Dunn Index, will be computed to assess the quality of generated clusters. Although the purpose is to predict stock market volatility in India given by historic volatility of NIFTY returns with the help of the predictors mentioned above, our study is an exploration of patterns in the data to understand whether volatility can be predicted at all.

Accordingly, the plan of the chapter is as follows. Section 3 explains the methodology of the study and a literature review is presented in Section 4. The variables are explained

in Section 5 and Section 6 presents the results. Some concluding observations are made in Section 7.

## Methodology

Clustering is the process of partitioning the data objects into segments of homogeneous data objects based on similarity of some features. Each segment is known as a cluster. Objects belonging to a particular cluster are similar to one another and dissimilar to objects belonging to other clusters. It is an unsupervised learning process as no prior information about the class of data objects is available. Meaningful knowledge can only be inferred once the given set of data points is grouped into different clusters. Mathematically in N-dimensional Euclidean space, the task of clustering is to partition a given set (S) of data points $\{x_1, x_2, x_3, \ldots\ldots, x_n\}$ into K clusters $\{C_1, C_2, C_3, \ldots\ldots..., C_n\}$ where the following conditions are satisfied:

$C_i \neq \emptyset$ for i=1,2,…., K (1)

$C_i \cap C_j = \emptyset$ for i=1,2,…, K; j=1,2,…, K and i≠j (2)

and

$\cup_{i=1}^{K} C_i = S$ (3)

Different clustering algorithms such as partitioning, divisive, density based and spectral clustering have been proposed and discussed throughout the literature. Similarly based on the nature of assignment of an object to a particular cluster, clustering techniques are classified as soft, fuzzy and probabilistic clustering. Some algorithms require number of clusters to be defined beforehand, while some others adjust the number of clusters based on some statistical measures respectively. To analyze the outcome of clustering or to assess the quality of formed clusters, broadly three different measures, internal, external and relative measures for clustering validations are usually applied. External measures

are supervised techniques that compare the outcome against some prior ground truth information or expert-specified knowhow. Whereas internal measures are completely unsupervised techniques which measure the goodness of results by determining how well the clusters are separated and how compact they are. The approach of relative measures is to compare different clusters obtained by different parameter setting of same algorithms. Brief descriptions of working principles of these algorithms are provided below.

### Kernel K-Means

It is a generalization of popular K-Means algorithm that overcomes the bottlenecks of the latter one. K-Means, a simple yet effective clustering tool, suffers if the data objects are not linearly separable. K-Means algorithm also fails to detect clusters which are not convex shaped. To overcome this obstacle Kernel K-Means algorithm projects data points of input space to a high dimensional feature space by applying nonlinear transformation functions (Kernel functions). Subsequently it follows the same principle of K-Means clustering algorithm in feature space to detect clusters. This algorithm initially generates a kernel matrix ($K_{ij}$) using equation

$$K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j) \quad (4)$$

where $x_i$, $x_j$ are data points to be clustered in input space. Usually a kernel function $K(x_i, x_j)$ is used to carry out the inner products in the feature space without explicitly defining transformation $\varphi$. Table 1 displays few well studied kernel functions as reported in literature.

**Table 1.** *Kernel Functions*

| | |
|---|---|
| Radial Basis Kernel | $K(x_i, x_j) = \exp\left(-\|x_i - x_j\|^2 / 2\sigma^2\right)$ |
| Polynomial Kernel | $K(x_i, x_j) = \left(x_i^T x_j + \gamma\right)^{\delta}$ |
| Sigmoid Kernel | $K(x_i, x_j) = \tanh\left(\gamma\left(x_i^T x_j\right) + \theta\right)$ |

**Source:** Authors' own construction

The outline of Kernel K-Means algorithm is illustrated below.

Step 1: Compute the Kernel matrix and initialize K cluster ($C_1, C_2, \ldots, C_k$) Centers arbitrarily.
Step 2: For each point $x_n$ and every cluster $C_i$ compute
$$|| \varphi(x_n) - m_i ||^2$$
Step 3: Find $c^*(x_n) = \text{argmin}\left(\|\varphi(x_n) - m_i\|^2\right)$
Step 4: Update clusters as $C_i = \{x_n | c^*(x_n)=i\}$
Step 5: Repeat steps 1 - 4 until convergence.

## Gaussian mixture model

It is a probabilistic clustering tool where the objective is to infer a set of probabilistic clusters which is most likely to generate the data set aimed to be clustered. If S be a set of m probabilistic clusters ($s_1, s_2, \ldots, s_m$) with probability density function ($f_1, f_2, \ldots, f_m$) and probabilities $w_1, w_2, \ldots, w_m$ respectively, then for any data point d, the probability that d is generated by cluster $s_i$ is given by $P(d|s_i) = w_i f_i(d)$. The probability that d is generated by the set S of clusters is computed as

$$P(d|S) = \sum_{i=1}^{m} w_i f_i(d) \quad (5)$$

If the data points are generated independently for data set, $D = (d_1, d_2, \ldots, d_n)$, then

$$P(D|S) = \prod_{j=1}^{n} P\left(d_j|S\right) = \prod_{j=1}^{n} \sum_{i=1}^{m} w_i f_i(d_i) \quad (6)$$

The objective of probabilistic model based clustering is to find a set of S probabilistic clusters such that P(D|S) is maximized. If the probability distribution functions are assumed to be Gaussian then the approach is known as Gaussian Mixture model. A multivariate Gaussian distribution function is characterized by the mean vector and covariance matrix. These parameters are estimated by Expectation Maximization algorithm.

In general if the data objects and parameters of m distribution are denoted by $D=\{d_1,d_2,\ldots,d_n\}$ and $\Theta=\{\Theta_1, \Theta_2,\ldots,\Theta_m\}$ then equation 5 may be expressed as

$$P(d_i|\Theta) = \sum_{i=1}^{m} w_i P_i(di|\Theta_i) \quad (7)$$

$P_i(d_i|\Theta_i)$ is the probability that di is generated from jth distribution using parameter $\Theta_i$. Equation 6 can be rewritten as

$$P(D|\Theta)= \prod_{j=1}^{n} \sum_{i=1}^{m} w_i P_i\left(d_j|\Theta_i\right) \quad (8)$$

For Gaussian Mixture Model, the objective is to estimate the parameters (mean vector and covariance matrix) that maximize equation 8.

Probability Distribution function of Gaussian distribution function is given by the following Formula

$$P(d_i|\Theta)= \frac{1}{(2\pi)^{l/2}\sqrt{|\sum|}} exp\left(-\frac{(d_i-\mu)^T\sum^{-1}(d_i-\mu)}{2}\right) \quad (9)$$

Where $\mu$ and $\Sigma$ are the mean and co-variance matrix of Gaussian and $l$ is the dimension of object $d_i$.

In Gaussian Mixture Model, the objective is to estimate the parameters (mean and covariance matrix) by Expectation Maximization (EM) algorithm that maximizes equation 10.

$$\Theta^* = \arg_{\Theta} \max P(D \mid \Theta) \quad (10)$$

Generally $\log P(D \mid \Theta)$ is maximized because of easier computations.

$$\log P(D \mid \Theta) = \log\left(\prod_{j=1}^{n} P(d_j \mid \Theta)\right) = \sum_{j=1}^{n} log\left(\sum_{i=1}^{m} w_i P_i(d_j \mid \Theta_i)\right) \quad (11)$$

An auxiliary objective function, Q is considered instead directly maximizing the log likelihood.

$$Q = \sum_{j=1}^{n} \sum_{i=1}^{m} \alpha_{ij} log\left[w_i P_i(d_j \mid \Theta_i)\right] \quad (12)$$

Where $\alpha_{ij}$ is the respective posteriori probabilities for individual class i.

$$\alpha_{ij} = \frac{w_i P_i(d_j \mid \Theta_i)}{\sum_{r=1}^{k} P_r(d_j \mid \Theta_r)} \quad (13)$$

$$\sum_{i=1}^{n} \alpha_{ij} = 1 \quad (14)$$

Maximizing equation ensures $P(D \mid \Theta)$ is maximized if performed by an EM algorithm. The steps of EM algorithm is given below

E-Step: Compute 'expected' classes of all data points for each class using Equation 7.

M-Step: Maximum likelihood given the data's class membership distributions is computed by the following equations.

$$W_i^{new} = \frac{1}{m} \sum_{j=1}^{m} \alpha_{ij} \quad (15)$$

$$\mu_i^{new} = \frac{\sum_{j=1}^{m} \alpha_{ij} d_j}{\sum_{j=1}^{m} \alpha_{ij}} \quad (16)$$

$$\sum_i^{new} = \frac{\sum_{j=1}^{m} \alpha_{ij} (d_j - \mu_i^{new})(d_j - \mu_i^{new})^T}{\sum_{j=1}^{m} \alpha_{ij}} \quad (17)$$

## Self-organizing map

Self-Organizing Maps (SOM) belong to nonlinear Artificial Neural Network models proposed by Kohonen (1990). It is an unsupervised learning algorithm mainly deployed to reduce dimensions of data set and to find homogenous groupings (clusters) among the data points. It basically attempts to visualize high dimensional data patterns onto one or two dimensional grid or lattice of units (neurons) adaptively in a topological ordered manner. This transformation tries to preserve topological relations, i.e., patterns which are similar in the input space will be mapped to units that are close in the output space as well, and vice-versa. The units are connected to adjacent ones by a neighborhood relation which is varied dynamically in the network training process. The number of neurons accounts for the accuracy and generalization capability of the SOM. All neurons compete for each input pattern; the neuron that is chosen for the input pattern wins it. Only the winning neuron is activated (winner-takes-all). The winning neuron updates itself and neighbor neurons to approximate the distribution of the patterns in the input dataset. After the adaptation process is complete, similar clusters will be close to each other (i.e., topological ordering of clusters). The SOM network organizes itself by competing representation of the samples. Neurons are also allowed to change themselves in hoping to win the next competition. This selection and learning process makes the weights to organize themselves into a map representing similarities. The three key steps to form self-organizing maps are known as completion phase (identifying the best matching or winning neuron), cooperation phase (determining the location of topological neighborhood) and synaptic adaptation phase (self-organized formation of feature maps). The SOM algorithm is summarized below:

1. Initialization: Randomly initialize the weight vectors $W_j(0)$, where $j = 1,2,\ldots\ldots,l$ and $l$ is the number of neurons in grid.
2. Sampling: Draw a sample training input vector x from the input space.
3. Similarity Matching: Find the best matching (winning) neuron i(x) at time step n by using minimum-distance criterion.

$$i(x) = \arg\min_j ||x(n) - w_j||, \quad j = 1,2,\ldots\ldots\ldots,l$$

4. Updating: Adjust the synaptic-weight vectors of all excited neurons

$$w_j(n+1) = w_j(n) + \varepsilon(n)\, h_{j,i(x)}(n)(x(n)-w_j(n)),$$

where $\varepsilon(n)$is learning rate and $h_{j,i(x)}(n)$ is the neighborhood function centered around i(x), the winning or best matching unit. In this study neighborhood function is computed as

$$h_{j,i(x)}(n) = \exp\left(-\frac{d_{j,i}^2}{2\sigma^2}\right)$$

Parameter $\sigma$ is the effective width of the topological neighborhood.
5. Continuation: Repeat step 2-4 until convergence.

Due to unavailability of ground truth information, we have opted for internal clustering validity index measures. Basically they evaluate a clustering by analyzing separation of and compactness of individual clusters. These indices sometimes are also applied to automatically determine the number of clusters. However, in this study instead of fixing the number of clusters, these measures are computed across a range of number of clusters as the objective is to infer the nature of stock market volatility. Silhouette Index (SI), Dunn Index (DI), Alternative Dunn index (ADI), Krzanowski–Lai index (KL) and Calinski–Harabasz index (CH) are examples of various internal validation measures which have been used frequently in different applications reported in literature. Here we have employed Silhouette Index (SI) and Dunn Index (DI) separately to assess the clustering results.

### Silhouette index (SI)

For a dataset D of n objects, if D is partitioned into k clusters, $C_1,\ldots,C_K$, Silhouette Index, s(i) for each object $i \in D$ is computed as

$$s(i) = \frac{b(i)-a(i)}{max\{a(i),b(i)\}}$$

Here $a(i)$ is the average distance between i and all other objects in the cluster in which i belongs whereas $b(i)$ is the minimum average distance from i to all clusters to which i does not belong. The Value of SI ranges between -1 and 1. A larger value indicates better quality clustering result.

### Dunn index (DI)

DI is a ratio between the minimal inter cluster distance to maximal intra cluster distance. The index is computed as

$$D = \frac{d_{min}}{d_{max}}$$

where $d_{min}$ represents the smallest distance between two objects from different clusters and $d_{max}$ denotes the largest distance of two objects from the same cluster. Larger value of DI implies better quality clusters.

## Literature review

Clustering is an active area of data mining research and many applications in the area of image and video processing, telecommunication churn management, stock market analysis, system biology, social network analysis and cellular manufacturing have been reported in the literature. Ozer (2001) utilized fuzzy clustering analysis for user segmentation of online music services. Nanda *et al.* (2010) adopted K-Means, Self-Organizing Maps and Fuzzy C-Means based clustering algorithm to classify Indian stocks in

different clusters and subsequently developed portfolios from these clusters. Kim & Ahn (2008) applied a Genetic Algorithm based K-Means clustering algorithm to develop recommender system for online shopping market. Siyal & Yu (2005) proposed a modified FCM algorithm for bias (also called intensity in-homogeneities) estimation and segmentation of MR (Magnetic resonance) images. Sun & Wing (2005) utilized K-Means algorithm to study the effect and implementation of different critical success factors for new product development in Hong Kong toy industry. Chattopadhyay *et al.* (2011) proposed a novel framework based on principal component analysis (PCA) and Self-Organizing Map (SOM) to carry out automatic cell formation in cellular manufacturing layouts.

Apart from applications based studies, significant amount of research work has been dedicated towards fundamental development of clustering methods. Maulik & Bandyopadhyay (2000) introduced genetic algorithm based clustering algorithm which displayed performance superiority over K-Means algorithm on artificial and real life data sets. Mitra *et al.* (2010) proposed a new clustering technique, Shadowed C-Means, integrating fundamental principles of fuzzy and rough clustering techniques. Later Mitra *et al.* (2011) utilized this algorithm for satellite image segmentation. Ju & Liu (2010) introduced fuzzy Gaussian mixture model (FGMM) based clustering hybridizing conventional Gaussian Mixture Model and Fuzzy set theory for faster convergence and tackling nonlinear data set. Hatamlou (2012) developed a new heuristic optimization based clustering technique, Black Hole algorithm, which outperformed several standard clustering methods. Chaira *et al.* (2011) proposed a new Intuitionistic Fuzzy C-Means algorithm defined on intuitionistic fuzzy set and successfully applied it to cluster CT scan brain images. There are many other clustering algorithms namely, Neural Gas, Artificial

Bee Colony Based Clustering technique (ABC), Gravitational search approach (GSA), Particle Swarm Optimization (PSO) based approach, Ant Colony Optimization (ACO) based technique, Chamelleon and DBSCAN that have been reported in literature.

Application of Decile Analysis in forecasting stock market volatility can be seen in McMillan (2004) and Datta Chaudhuri & Sheth (2014). The literature on volatility prediction by the ARCH/GARCH method includes papers by Das & Bhattacharya (2014), Karolyi (1995), Kumar & Mukhopadhyay (2007), Angela (2000), and Padhi & Logesh (2012). Datta Chaudhuri & Ghosh (2015), deployed Artificial Neural Network based framework for prediction of stock market volatility in the Indian stock market through volatility of NIFTY returns and volatility of gold returns.

## The variables

For our analysis we have considered daily data of nine variables namely volatility of NIFTY returns (NIFTYSDR), volatility of gold returns (GOLDSDR), India VIX, CBOE VIX, volatility of crude oil returns (CRUDESDR), volatility of DJIA returns (DJIASDR), volatility of DAX returns (DAXSDR), volatility of Hang Seng returns (HANGSDR) and volatility of Nikkei returns (NIKKEISDR) for the years 2013 and 2014. In the analysis there are no inputs or outputs. All the variables are considered together to identify clusters. However, the implicit reason for choosing the variables is that there does exist some association between them and hence do play a role in explaining historic volatility. Figures 2 and 3 provide examples of two such long term associations

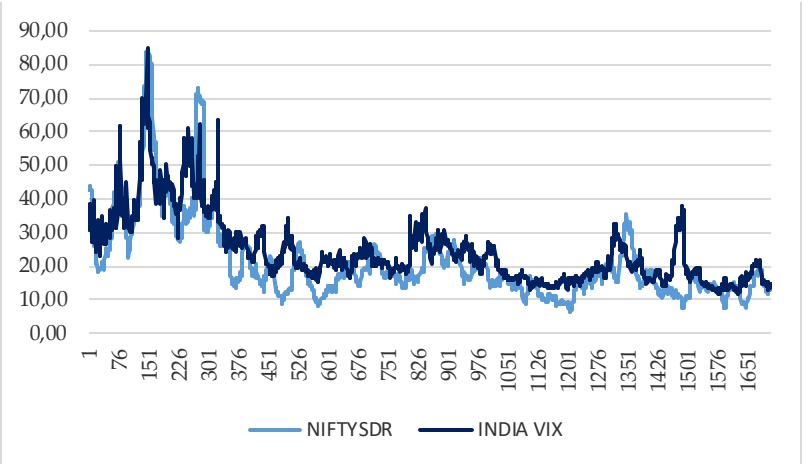


**Figure 2.** *INDIAVIX and NIFTYSDR for the period 3.3.2008 to 10.4.2015*

**Source:** Authors' own construction

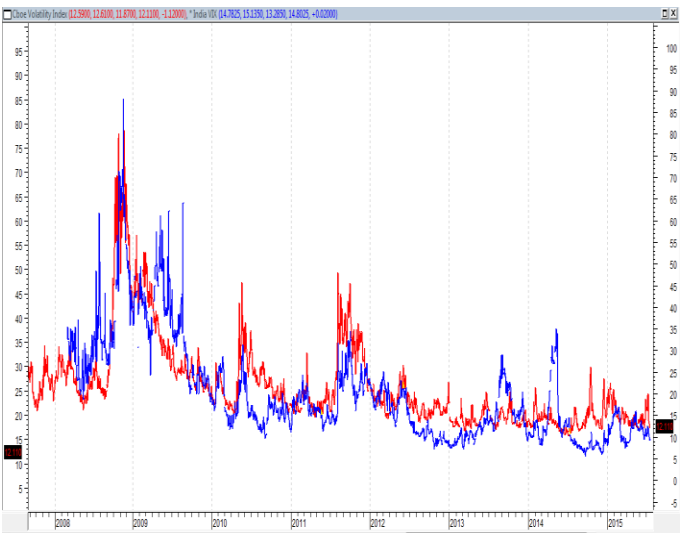

**Figure 3.** *INDIA VIX AND CBOE VIX FOR THE PERIOD 2008 – 2015 (June)*

**Source:** Metastock

Figure 2 indicates that, over a fairly long period, historic volatility and implied volatility do move together. So

considering INDIA VIX as a predictor of NIFTYSDR is alright. Further, it may be observed from Figure 3 that expected volatility in the US seems to go hand in hand with expected volatility in India. That is, global uncertainties affect US implied volatility, which in turn affects implied volatility in India. To allow for external shocks, as India is a large importer of crude oil, we consider CRUDESDR in the analysis. In the recent past, political instability in the Middle East and related regions has impacted the expected availability of oil and has resulted in stock market instability in India. Global instability, both in the western and the eastern world has been incorporated through DJIASDR, DAXSDR, HANGSDR and NIKKEISDR.

## Results and analysis

Tables 2 to 4 present the obtained DI values of the clusters generated by the three algorithms for different combinations of features and number of clusters.

**Table 2.** *DI values of clustering result generated by Kernel K-Means algorithm*

| | | No. of Features | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* |
| No. of Clusters | 2 | 0.0303 | 0.0573 | 0.0443 | 0.0447 | 0.0499 | 0.0926 | 0.0979 | 0.0976 |
| | 3 | 0.0322 | 0.102 | 0.107 | 0.0509 | 0.0697 | 0.0499 | 0.0859 | 0.0519 |
| | 4 | 0.0282 | 0.0987 | 0.0497 | 0.0988 | 0.0561 | 0.0977 | 0.1248 | 0.0608 |
| | 5 | 0.0813 | 0.0947 | 0.0473 | 0.099 | 0.1363 | 0.1361 | 0.1454 | 0.1451 |
| | 6 | 0.0841 | 0.0598 | 0.0476 | 0.0831 | 0.0913 | 0.0795 | 0.1251 | 0.1222 |
| | 7 | 0.0327 | 0.0845 | 0.0954 | 0.0831 | 0.1632 | 0.1686 | 0.0985 | 0.0718 |
| | 8 | 0.0841 | 0.0889 | 0.0425 | 0.0736 | 0.126 | 0.1419 | 0.1108 | 0.1102 |
| | 9 | 0.0813 | 0.0928 | 0.0918 | 0.0898 | 0.126 | 0.1563 | 0.1108 | 0.1102 |
| | 10 | 0.0867 | 0.1561 | 0.0993 | 0.0909 | 0.126 | 0.1285 | 0.1228 | 0.1102 |
| | 11 | 0.0661 | 0.1059 | 0.0993 | 0.1687 | 0.133 | 0.1285 | 0.1234 | 0.1186 |

**Table 3.** *DI values of clustering result generated by Self-Organizing Map*

| | | No. of Features | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* |
| No. of Clusters | 2 | 0.0237 | 0.0515 | 0.0298 | 0.0198 | 0.0571 | 0.0499 | 0.0979 | 0.0976 |
| | 3 | 0.0339 | 0.102 | 0.039 | 0.1309 | 0.0392 | 0.0344 | 0.0764 | 0.0875 |
| | 4 | 0.0282 | 0.0987 | 0.1039 | 0.0596 | 0.0919 | 0.0557 | 0.0723 | 0.0671 |
| | 5 | 0.0476 | 0.0348 | 0.0473 | 0.099 | 0.0754 | 0.0717 | 0.0817 | 0.0987 |
| | 6 | 0.0421 | 0.0743 | 0.0482 | 0.0831 | 0.1247 | 0.1279 | 0.0913 | 0.081 |
| | 7 | 0.0501 | 0.0606 | 0.0461 | 0.0681 | 0.1632 | 0.1686 | 0.0985 | 0.1518 |
| | 8 | 0.0379 | 0.0603 | 0.0435 | 0.0987 | 0.0959 | 0.1119 | 0.1108 | 0.1102 |
| | 9 | 0.0545 | 0.0889 | 0.0918 | 0.086 | 0.126 | 0.1122 | 0.0669 | 0.109 |
| | 10 | 0.0545 | 0.1427 | 0.096 | 0.0718 | 0.1423 | 0.1113 | 0.0908 | 0.0907 |
| | 11 | 0.0578 | 0.1463 | 0.0994 | 0.0909 | 0.0723 | 0.1807 | 0.0829 | 0.0907 |

**Table 4.** *DI values of clustering result generated by Gaussian Mixture Mode*

| | | No. of Features | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* |
| No. of Clusters | 2 | 0.0103 | 0.1456 | 0.0473 | 0.0317 | 0.0313 | 0.1436 | 0.0675 | 0.1816 |
| | 3 | 0.0048 | 0.0749 | 0.0359 | 0.228 | 0.0587 | 0.1436 | 0.0744 | 0.0541 |
| | 4 | 0.0272 | 0.0592 | 0.0208 | 0.2135 | 0.0188 | 0.0495 | 0.0744 | 0.1152 |
| | 5 | 0.0379 | 0.0377 | 0.091 | 0.2796 | 0.0722 | 0.0729 | 0.0858 | 0.1152 |
| | 6 | 0.048 | 0.0288 | 0.0526 | 0.3079 | 0.0837 | 0.0985 | 0.0858 | 0.1428 |
| | 7 | 0.0377 | 0.0403 | 0.0441 | 0.3581 | 0.0837 | 0.0985 | 0.0925 | 0.0687 |
| | 8 | 0.0449 | 0.054 | 0.0701 | 0.3482 | 0.1095 | 0.0985 | 0.0925 | 0.0687 |
| | 9 | 0.0209 | 0.0362 | 0.0309 | 0.3183 | 0.0954 | 0.0985 | 0.1033 | 0.1307 |
| | 10 | 0.0288 | 0.0478 | 0.0475 | 0.3533 | 0.0954 | 0.0985 | 0.1318 | 0.0903 |
| | 11 | 0.0499 | 0.0387 | 0.1036 | 0.3808 | 0.112 | 0.1748 | 0.1593 | 0.111 |

In Table 2, the maximum DI value of 0.1686 corresponds to 5 features (India VIX, NIFTYSDR, CBOE VIX, volatility of crude oil returns, volatility of DJIA returns) and 7 clusters. Similarly, in Tables 3 and 4, maximum DI values correspond to 7 features (India VIX, NIFTYSDR, CBOE VIX, volatility of crude oil returns, volatility of DJIA returns, volatility of DAX returns, volatility of Hang Seng returns) 11 clusters and 5 features (India VIX, NIFTYSDR, CBOE VIX, volatility of

crude oil returns, volatility of DJIA returns) 11 clusters respectively. For better understanding, following figures map the relationship between number of features and number of clusters. Five common features present in all three experiments are India VIX, NIFTYSDR, CBOE VIX, volatility of crude oil returns and volatility of DJIA returns.

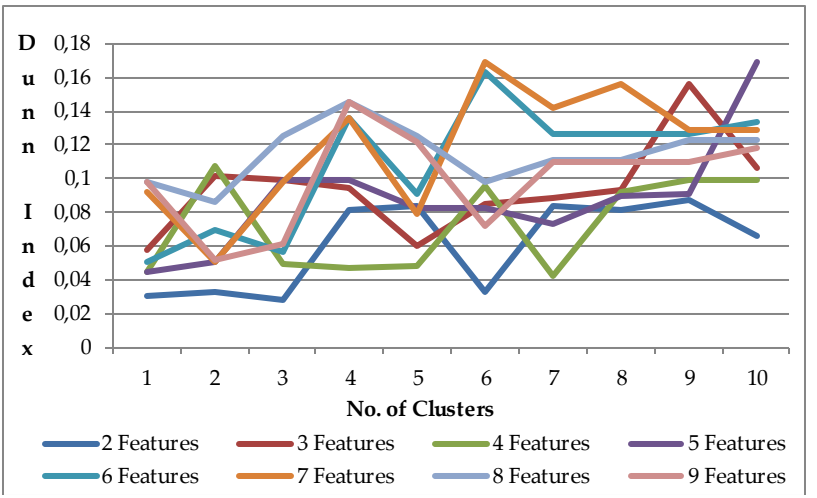


**Figure 4.** *DI values of clustering/features generated by Kernel K-Means algorithm*

As the Figure 1 contains several spikes (both in positive and negative direction corresponding to local maxima and minima) it is hard to determine whether incremental increase in number clusters result in good or bad quality clusters. However, it may be broadly inferred that large number of features (6-9) produces better quality segmentation in compared to smaller number of features (1-3). Figure 2 justifies the claim as well. Figure 3 clearly identifies that usage of 5 features (India VIX, NIFTYSDR, CBOE VIX, volatility of crude oil returns, volatility of DJIA returns) yields superior cluster quality than other combinations.

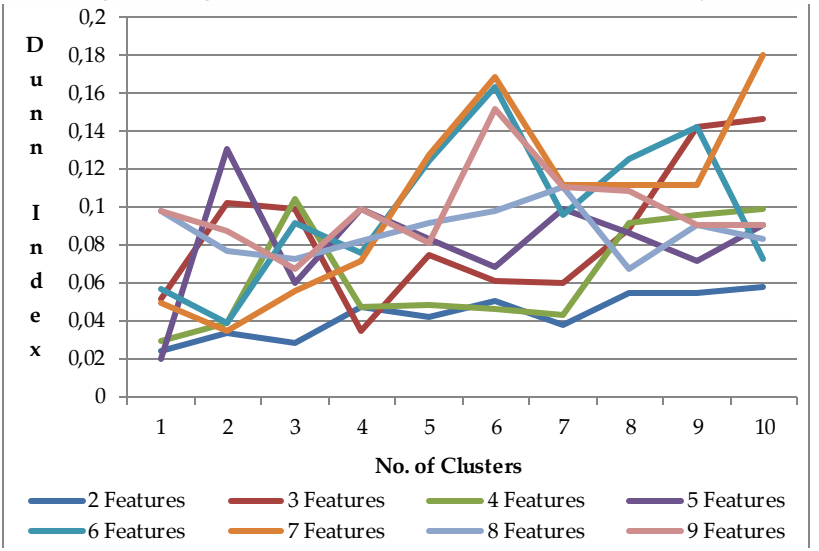


**Figure 5.** *DI values of clustering/features generated by Gaussian Mixture Model*

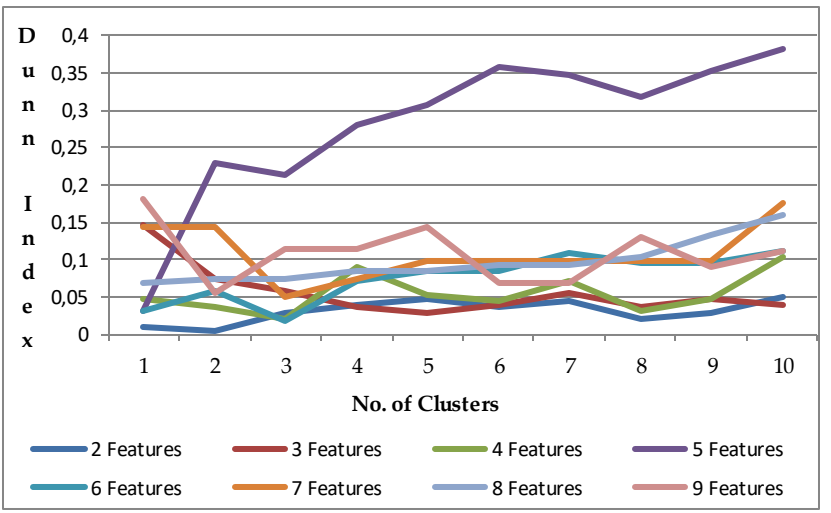


**Figure 6.** *DI values of clustering/features generated by Self Organizing Maps*

Same clustering algorithms are applied on the same data set to calculate Silhouette Index values. Results are summarized in tables 5-7.

**Table 5.** *SI values of clustering result generated by Kernel K-Means algorithm*

| | | No. of Features | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* |
| No. of Clusters | 2 | 0.5914 | 0.4809 | 0.4062 | 0.3786 | 0.356 | 0.3446 | 0.3447 | 0.3372 |
| | 3 | 0.5994 | 0.5147 | 0.4525 | 0.4278 | 0.3912 | 0.3805 | 0.3331 | 0.3689 |
| | 4 | 0.5527 | 0.4941 | 0.4183 | 0.4592 | 0.4268 | 0.4071 | 0.3811 | 0.3579 |
| | 5 | 0.5659 | 0.4823 | 0.4601 | 0.4568 | 0.4606 | 0.4419 | 0.4168 | 0.4091 |
| | 6 | 0.5121 | 0.4734 | 0.4461 | 0.4614 | 0.4868 | 0.4617 | 0.437 | 0.4284 |
| | 7 | 0.4836 | 0.4156 | 0.4631 | 0.466 | 0.4731 | 0.4551 | 0.4395 | 0.4173 |
| | 8 | 0.5266 | 0.4736 | 0.4494 | 0.4323 | 0.4669 | 0.4464 | 0.4446 | 0.4376 |
| | 9 | 0.5129 | 0.4774 | 0.4575 | 0.4248 | 0.4169 | 0.4476 | 0.45 | 0.4232 |
| | 10 | 0.4965 | 0.4681 | 0.4349 | 0.4157 | 0.4154 | 0.4481 | 0.4515 | 0.4351 |
| | 11 | 0.4966 | 0.4378 | 0.4285 | 0.4207 | 0.4076 | 0.4383 | 0.4469 | 0.4326 |

**Table 6.** *SI values of clustering result generated by Self-Organizing Map*

| | | No. of Features | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* |
| No. of Clusters | 2 | 0.4733 | 0.4741 | 0.4043 | 0.4043 | 0.3568 | 0.345 | 0.3462 | 0.3372 |
| | 3 | 0.5932 | 0.5147 | 0.4474 | 0.4257 | 0.3369 | 0.345 | 0.371 | 0.3691 |
| | 4 | 0.5527 | 0.4941 | 0.4606 | 0.4577 | 0.4233 | 0.3791 | 0.3973 | 0.3944 |
| | 5 | 0.5246 | 0.45 | 0.4601 | 0.4569 | 0.4458 | 0.4109 | 0.397 | 0.4129 |
| | 6 | 0.4868 | 0.466 | 0.4155 | 0.4614 | 0.4297 | 0.4234 | 0.3883 | 0.372 |
| | 7 | 0.4616 | 0.4582 | 0.4431 | 0.4309 | 0.472 | 0.4134 | 0.4381 | 0.4309 |
| | 8 | 0.4525 | 0.4274 | 0.4431 | 0.4004 | 0.4643 | 0.4551 | 0.4446 | 0.4376 |
| | 9 | 0.4875 | 0.4401 | 0.4575 | 0.4368 | 0.4637 | 0.4505 | 0.4299 | 0.4201 |
| | 10 | 0.4778 | 0.4279 | 0.432 | 0.418 | 0.4382 | 0.3641 | 0.4366 | 0.409 |
| | 11 | 0.4562 | 0.4221 | 0.4733 | 0.4111 | 0.4334 | 0.4411 | 0.4019 | 0.406 |

**Table 7.** *SI values of clustering result generated by Gaussians Mixture Model*

| | | No. of Features | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* |
| No. of Clusters | 2 | 0.4733 | 0.4741 | 0.4043 | 0.4043 | 0.3568 | 0.345 | 0.3462 | 0.3372 |
| | 3 | 0.5932 | 0.5147 | 0.4474 | 0.4257 | 0.3369 | 0.345 | 0.371 | 0.3691 |
| | 4 | 0.5527 | 0.4941 | 0.4606 | 0.4577 | 0.4233 | 0.3791 | 0.3973 | 0.3944 |
| | 5 | 0.5246 | 0.45 | 0.4601 | 0.4569 | 0.4458 | 0.4109 | 0.397 | 0.4129 |
| | 6 | 0.4868 | 0.466 | 0.4155 | 0.4614 | 0.4297 | 0.4234 | 0.3883 | 0.372 |
| | 7 | 0.4616 | 0.4582 | 0.4431 | 0.4309 | 0.472 | 0.4134 | 0.4381 | 0.4309 |
| | 8 | 0.4525 | 0.4274 | 0.4431 | 0.4004 | 0.4643 | 0.4551 | 0.4446 | 0.4376 |
| | 9 | 0.4875 | 0.4401 | 0.4575 | 0.4368 | 0.4637 | 0.4505 | 0.4299 | 0.4201 |
| | 10 | 0.4778 | 0.4279 | 0.432 | 0.418 | 0.4382 | 0.3641 | 0.4366 | 0.409 |
| | 11 | 0.4562 | 0.4221 | 0.4733 | 0.4111 | 0.4334 | 0.4411 | 0.4019 | 0.406 |

Maximum SI values of table 5, 6 and 7 correspond to 2 features (India VIX and NIFTYSDR) and 3 clusters (0.5994), 2 features (India VIX and NIFTYSDR) and 3 clusters (0.5932), 2 features (India VIX and NIFTYSDR) and 3 clusters (0.5932) respectively. Unlike the pattern observed in DI values, here it is quite evident that increase in number of clusters does not improve the quality of clusters. It also indicates that addition of extra features fails to enhance clusters quality significantly as well. The results are depicted in following figures.

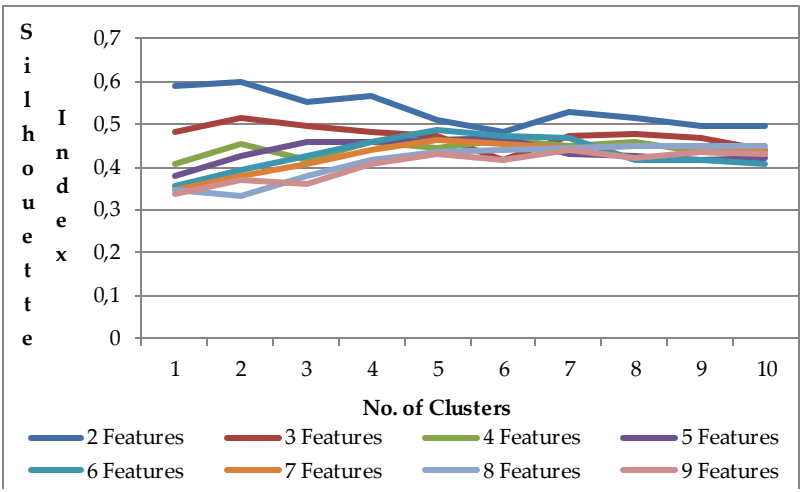


**Figure 7.** *SI Index obtained by Kernel K-Means technique*

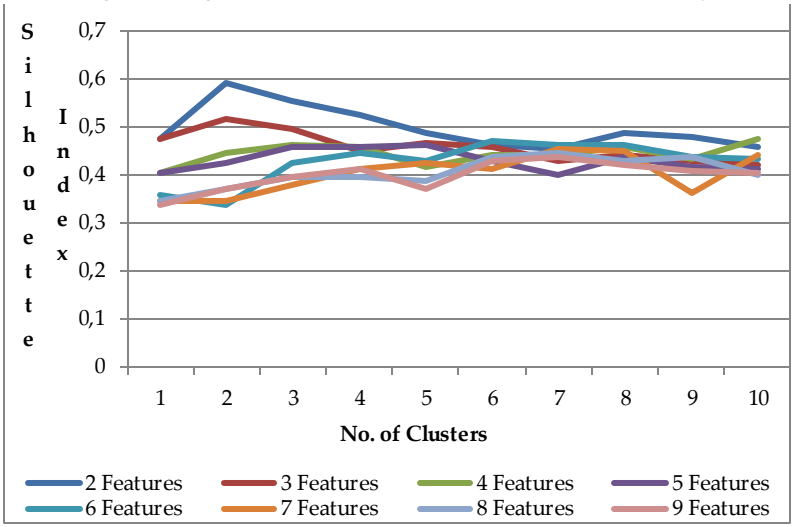


**Figure 8.** *SI Index obtained by Self-Organizing Map technique*

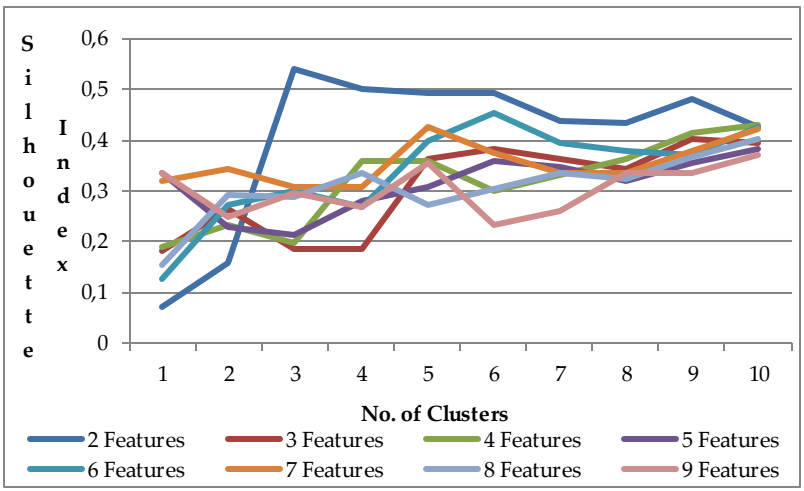


**Figure 9.** *SI Index obtained by Gaussian Mixture Model*

## Concluding remarks

The purpose of this chapter was to demonstrate that volatility prediction in stock markets has to be preceded by a study of the number of predictors and the number of clusters. The data on historic volatility may not be homogenous and the presence of many clusters would

validate that. If there are too many clusters then it implies that volatility is random and would be difficult to predict. Further, the choice of the predictors has to be mapped with the number of clusters. Too many predictors with large number of clusters over a long time series data may not yield efficient results. Our study for two years for the Indian stock market reveals that of the variables chosen, seven predictors over five to six clusters gave optimum results. This implies that, given the time span as defined by a cluster, one can predict volatility with the help of the predictors. For data spanning across clusters, prediction may not be desirable. Diagrams of the Silhouette Index for the algorithms indicate that the data in the sample can at most be broken into three clusters. This implies that three broad distinct associations were seen among the variables chosen, and within the clusters forecasting is possible.

## References

Angela, N. (2000). Volatility spillover effects from Japan and the US to Pacific-Basin, *Journal of International Money and Finance*, 19(2), 207-233. doi. 10.1016/S0261-5606(00)00006-1

Chaira, T. (2011). A novel intuitionistic fuzzy C means clustering algorithm and its application to medical images, *Applied Soft Computing*, 11(2), 1711-1717. doi. 10.1016/j.asoc.2010.05.005

Chattopadhyay, M., Dan, P.K., & Mazumdar, S. (2011). Principal component analysis and self-organizing map for visual clustering of machine-part cell formation in cellular manufacturing system, *Systems Research Forum*, 5(1), 25-51. doi. 10.1142/S179396661100028X

Das, S., & Bhattacharya, B. (2014). Global financial crisis and pattern of return and volatility spill-over from the stock markets of USA and Japan on the Indian stock market: An application of EGARCH model, *CBS Journal of Management Practices*, 1(1), 1-18.

Datta Chaudhuri, T., & Kinjal, S. (2014). Forecasting volatility, volatility trading and decomposition by Greeks, *CBS Journal of Management Practices*, 1(1), 59-70.

Datta Chaudhuri, T., & Ghosh, I. (2015). Forecasting volatility in Indian stock market using artificial neural network with multiple inputs and outputs, *International Journal of Computer Applications*, 120(8), 7-15. doi. 10.5120/21245-4034

Hatamlou, A. (2012). Black hole: A new heuristic optimization approach for data clustering, *Information Sciences*, 222, 175-184. doi. 10.1016/j.ins.2012.08.023

Ju, Z., & Liu, H. (2012). Fuzzy Gaussian mixture models, *Pattern Recognition*, 45(3), 1146–1158. doi. 10.1016/j.patcog.2011.08.028

Karloyi, G.A. (1995). A multivariate GARCH model of international transmissions of stock returns and volatility: The case of United States and Canada, *Journal of Business and Economic Statistics*, 31(1), 11-25.

Kim, K.-J., & Ahn, H. (2008). A recommender system using GA K-means clustering in an online shopping market. *Expert Systems with Applications*, 34(2), 1200-1209. doi. 10.1016/j.eswa.2006.12.025

Kumar, K.K., & Mukhopadhyay, C. (2002). Volatility spillovers from US to Indian stock market: A comparison of GARCH models, *ICFAI Journal of Financial Economics*, 5(4), 7-30.

Maulik, U., & Bandyopadhyay, S. (2000). Genetic algorithm-based clustering technique, *Pattern Recognition*, 33(9), 1455-1465. doi. 10.1016/S0031-3203(99)00137-5

McMillan, L.G. (2004). *McMillan on Options*, John Wiley & Sons, Inc., Hoboken, New Jersey.

Mitra, S., Pedrycz, W., & Barman, B. (2010). Shadowed c-means: Integrating fuzzy and rough clustering, *Pattern Recognition*, 43(4), 1282–1291. doi. 10.1016/j.patcog.2009.09.029

Mitra, S., & Kundu, P.P. (2011). Satellite image segmentation with Shadowed C-Means, *Information Sciences*, 181(17), 3601-3613. doi. 10.1016/j.ins.2011.04.027

Nanda, S.R., Mahanty, B., & Tiwari, M.K. (2010). Clustering Indian stock market data for portfolio management, *Expert Systems with Applications*, 37(12), 8793–8798. doi. 10.1016/j.eswa.2010.06.026

Ozer, M. (2001). User segmentation of online music services using fuzzy clustering, *Omega*, 29(2), 193-206. doi. **10.1016/S0305-0483(00)00042-6**

Padhi, P., & Lagesh, M.A. (2012). Volatility spillover and time varying correlation among the Indian and US Stock markets, *Journal of Quantitative Economics*, 10(2). 78-90.

Sinha, P., & Sinha, G. (2010). Volatility spillover in India, USA and Japan investigation of recession effects. [Retrieved from].

Siyal, M.Y., & Yu, L. (2005). An intelligent modified fuzzy c-means based algorithm for bias estimation and segmentation of brain MRI, *Pattern Recognition Letter*, 26(13), 2052–2062. doi. 10.1016/j.patrec.2005.03.019

Sun, H., & Wing, W.C. (2005) Critical success factors for new product development in the Hong Kong toy industry, *Technovation*, 25, 293–303. doi. 10.1016/S0166-4972(03)00097-X

**For Cited this Chapter:**

**Chaudhuri, T.D., & Ghosh, I. (2020). Using clustering method to understand Indian stock market volatility, *In* T.D. Chaudhuri (Ed.), *Overview of the Indian Economy: Empirical analysis of financial markets, exchange rate movements and industrial development* (pp.88-112), KSP Books: Istanbul.**

ISBN: 978-605-7736-93-2 (e-Book)
KSP Books 2020
© KSP Books 2020

**Copyrights**
Copyright for this Book is retained by the author(s), with first publication rights granted to the Book. This is an open-access Book distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by-nc/4.0 ).

# 5 Forecasting volatility in Indian stock market using artificial neural network with multiple inputs and outputs

*By*
Tamal DATTA CHAUDHURI
& Indranil GHOSH

## Introduction

Volatility in stock markets evokes varying responses from market participants. While some perceive it as opportunity to make money, others perceive it as a threat and start unwinding their positions. While affecting portfolio choice, changes in stock market volatility also gives some idea about the current economic state. In today's globalized environment, increased volatility reflects global uncertainty. Volatility in the stock market as a whole can be due to macroeconomic factors, both internal and external. Examples could be oil price shocks, or increase in rates of interest in the US, or domestic elections. Volatility in individual stocks, on the other hand, can be due to perceived growth prospects of the company or the sector. It could also be triggered by company specific news or policy announcements.

The effects of stock market volatility can be sixfold. First, it enhances the profit making opportunities from intraday tradingfor spot market traders. Second, it leads to portfolio rebalancing by fund managers. Third, it increases volatility

trading in the options market. Fourth, it increases hedging activity in financial markets. Fifth, it does influence policy makers in taking hard decisions as their actions can cause loss of wealth to retail holders. Sixth, it affects capital formation, as volatile markets are not conducive for fresh equity issues in the market.

While the effects of unanticipated announcements by companies, or external macroeconomic events like sovereign defaults, or economy wide policy changes on market volatility cannot be estimated, under normal market conditions, the Black and Scholes options pricing model provides a framework to estimate future volatility. This is denoted by "implied volatility", volatility that is expected to prevail in the near future as implied by the option price. In the spot market, if the expectation is that spot prices are going to fall, players rush to the options market to hedge their positions thus increasing implied volatility.

In the options pricing formula, the options price C

$$C = f(S, K, t, \sigma, r)$$

where S is the spot price of the underlying, K is the strike price, t is the time to expiry, $\sigma$ is volatility and r is the rate of interest. In the options market, the players cannot influence S, t or r. K they have to choose themselves. The only two variables that remain are C and $\sigma$. If we substitute the value of historic volatility in place of $\sigma$, then we will solve for the theoretical options price. If we plug in the value of the actual traded price of the options contract, then we will solve for implied volatility. The latter is an estimate of the actual volatility that is expected to prevail in the next three to four weeks. Thus, actual volatility and implied volatility should move together. The various possible movements between historic volatility and implied volatility has been described in detail in Passarelli (2008).

In todays globalized environment, with increased financial integration and also enhanced trade in goods and services, volatility in one country spreads to other countries almost immediately. In India, where foreign institutional investors (FIIs) are large players in the stock market, their fund allocation is shaped by macroeconomic conditions in other economies. Thus any macroeconomic event in any part of the world, causes reallocation of FII funds, leading to volatility in Indian stock markets.

Generally, when stock market becomes volatile, there is a tendency for gold prices to rise. It is considered to be a safe asset and hence there is a tendency to substitute stocks with gold. Thus volatility in gold prices is also a reflection of volatility in stock markets.

The purpose of this chapter is to develop a framework for forecasting volatility in the Indian stock market.

## Objective of the chapter

The chapter proposes an Artificial Neural Network (ANN) framework for forecasting volatility in the Indian stock market. The model has volatility of NIFTY returns and volatility of gold returns as the two outputs. It has India VIX, CBOE VIX, volatility of crude oil returns, volatility of DJIA returns, volatility of DAX returns, volatility of Hang Seng returns and volatility of Nikkei returns as the seven inputs. The objective is to capture the effects of both external and internal shocks on spot market volatility. The advantage of using the ANN framework is that it does not presuppose any linearity in the relationship between the inputs and outputs. Further, it allows for interaction and feedback between the inputs. We do not use lagged values of the outputs as inputs to avoid time dependency and we model external shocks through crude oil market volatility as well as volatility in other financial markets. Internal shocks are assumed to be represented through movements in India VIX.

Accordingly, the plan of the chapter is as follows. The ANN framework for our study is described in Section III. A literature survey is presented in Section IV. The choices of variables are discussed in Section V. The data and the results of the study are discussed in Section VI and Section VII concludes the chapter.

## Methodology

Artificial Neural Networks (ANN) are effective machine learning tools that mimic the working nature of human brain in order to identify the associative pattern between a set of inputs and outputs. Human brain is a massively interconnected structure of around $10^{10}$ number of basic processing units known as neurons. Similar to this architecture, in ANN, neurons are structured and connected in a hierarchical manner. A distinct input layer and output layer are interlinked (artificial synapses) through a single or multiple hidden layer(s).

Figure 1 depicts a typical ANN architecture with five inputs, one hidden layer and one output.

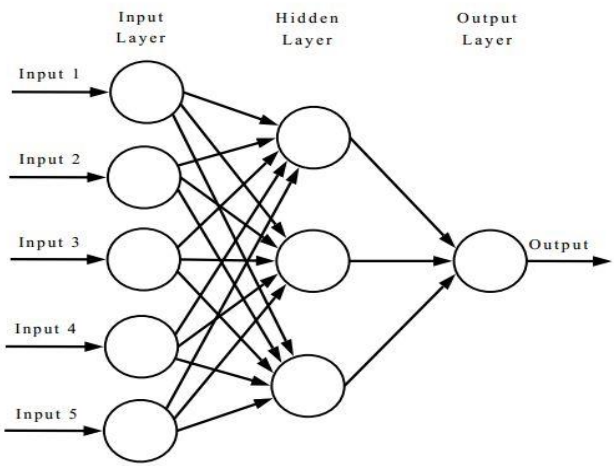


**Figure 1.** *A simple ANN model*

Strength of each connection between any two neurons is represented by numeric weight value. These weight values actually correspond to the decision boundary obtained by ANN classifier. When a given set of input and output values of variables under study are presented to a Neural Network as training dataset, weight values are estimated via different learning algorithms. Once the estimated values are stabilized after validation, trained ANN is tested against a test data set to evaluate its predictive power. Each input signal ($x_i$) is associated with a weight ($w_i$). The overall input I to the processing unit is a function of all weighted inputs given by

$$I = f\left(\sum x_i \times w_i\right) \quad (1)$$

The activation state of the processing unit (A) at any time is a function (usually nonlinear) of I

$$A = g(I) \quad (2)$$

The output Yfrom the processing unit is determined by the transfer function h

$$Y = h(A) = h(g(I)) = h(g(f(\sum x_i \times w_i)$$
$$= \Theta(\sum x_i \times w_i) \quad (3)$$

An ANNis said to "learn" mapping for a function or a process. Since the topology, the activation function A, and the transfer function h are normally fixed at the time the network is constructed, the only adjustable parameters are the weights $w_i$. Learning means changing the weights adaptively to meet some criterion based on the signals from the output units (nodes). A common training algorithm for ANN is back propagation (a steepest gradient descent method). It minimizes the sum of the squares of the differences between vectors Y and $Y_d$ .i.e.,

$$E = \frac{1}{2}(Y - Y_d)^T (Y - Y_d) \quad (4)$$

Where Y represents a vector of outputs of all the output nodes, $Y_d$is a vector of desired outputs, and superscript T stands for standard transpose operation. Many types of ANN models have been proposed during the last two decades to map inputs to outputs. Among them, layered ANN'strained by a back-propagation learning algorithm form the basis for the most common practical applications. Weight and bias matrix associated with the inputs are adjusted/updated by usingsome learning rule or training algorithm which is non-linear, multi-variable optimization (minimization) of error function. Based on the general relations in (1) through (3), the outputs from the input layer to the hidden layer and the outputs from the hidden layer to the output layer of the network are, respectively,

$$Z = \Theta_z(Z\,\Omega_z) \quad (5)$$

and

$$Y = \Theta_y(Z\,\Omega_y) \quad (6)$$

Where $\Theta_z$, and $\Theta_y$, are usually sigmoid functions which can be described by the following expression

$$O_j = 1/(1+\exp(i_j)) \quad (7)$$

Where $O_j$ is the output of node j and $i_j$ is the net-input of node j.

Due to its efficacy in parallel processing to mine complex nonlinear pattern, it has garnered a lot of attention in pattern recognition literature. Ability to operate in nonparametric environment has given it competitive edge over traditional statistical tools such as regression analysis. As a result of

which many variants of ANN model and different learning algorithms have been evolved. It has been highly successful both in predicting the state of categorical variable(s) and forecasting the outcome of continuous variable(s) as cited in literature. Complex real world problems such as prediction of financial health, bankruptcy prediction, stock index return analysis, credit default analysis, manufacturing assembly line balancing, PID controller monitoring, Enterprise Resource Planning performance analysis, etc. have been analyzed using the ANN framework and this is discussed in a subsequent section.

In this study Multilayer Feed Forward Network and Cascade Feed Forward Network has been adopted as ANN models and nine backpropagation algorithms namely Levenberg-Marquardt (LM), Scaled Conjugate Gradient (SCGA), Conjugate Gradient with Powell-Beale Restarts (CGB), BFGS Quasi-Newton (BFG), Resilient Backpropagation (RP), Fletcher-Powell Conjugate Gradient (CGF), Polak-Ribiére Conjugate Gradient (CGP), One Step Secant (OSS) and Variable Learning Rate Backpropagation (GDX) have been deployed for training purpose. Entire simulation process has been carried out using Neural Network toolbox of Matlab 7.9.0.

## Literature review

In this section, we present the two strands of the literature on which this chapter is based. The first is application of ANN in various areas of research which reflects the wide range applicability of this tool of analysis. The second is research papers on forecasting volatility in stock markets including those which have applied ANN as a tool of analysis.

Walczak & Sincich (1999) made a comparative analysis of neural network and logistic regression in student profile selection for university enrolments and the results showed

that ANN outperformed logistic regression. Ling & Liu (2004) investigated the critical success factors of design-build projects in Singapore through ANN based modelling where eleven success measures and sixtyfive factors were analyzed. Karnik *et al.* (2008) utilizedmultilayer feed forward ANN trained by backpropagation algorithm to model and critically examine the impact of drilling process parameters on the delamination factor. Pal *et al.* (2008) designed a multilayer ANN model to estimate the tensile stress of welded plates and compared the results with multiple regression analysis. Rouhani & Ravasan (2012) investigated the relationship between organizational factors and post Enterprise Resource Planning system implementation success using a novel Neural Network framework. Ndalimanetal (2012) proposed an ANN model with multi-layer perception neural architecture for the prediction of SR on first commenced Ti-15-3 alloy in electrical discharge machining (EDM) process. Zhao *et al.* (2015) utilized wavelet neural network and proposed a variable step size updating learning algorithm for parameter tuning operation of PID controller. Ramasamy *et al.* (2015) attempted to predict wind speeds of different locations (Bilaspur, Chamba, Kangra, Kinnaur, Kullu, Keylong, Mandi, Shimla, Sirmaur, Solan and Una location) in the Western Himalayan Indian state of Himachal Pradesh adopting ANN based framework. Ghiassi *et al.* (2015) applied dynamic artificial neural network to forecast movie revenues during the pre-production period in USA using MPAA rating, sequel, number of screens, production budgets, pre-release advertising expenditures, runtime &seasonality as predictor variables. Oko *et al.* (2015) presented a dynamic model of the drum-boiler to predict drum pressure and level in coal-fired subcritical power plant using NARX neural networks. Aish *et al.* (2015) incorporated Multilayer perceptron (MLP) and radial basis function (RBF)

neural networks as prediction tool to forecast reverse osmosis desalination plant's performanceinthe GazaStrip.

In the second strand of the literature, Rather *et al.* (2015) employed to two linear models namely auto regressive moving average and exponential smoothing, and recurrent neural network as nonlinear model to predict returns of six stocks (TCS, BHEL, Wipro, Axis Bank, Maruti& Tata Steel) using training dataset from National Stock Exchange of India (NSE). Results showed the supremacy of neural model over the linear models. Further, authors proposed a hybrid prediction model that use the results of individual prediction models and tested the effectiveness of it in estimating returns from twenty five stocks belonging to different industrial sectors.

Adhikary (2015) presented an ANN based ensemble prediction framework for time series forecasting problems. Malliaris & Salchenberger (1996) employed Elman's recurrent neural network and ARIMA model in forecasting copper spot prices using New York Commodity Exchange (COMEX) data. The study reports that the neural model outperform ARIMA model in terms of forecasting accuracy.

Malhotra (2012) attempted to examine the impact of stock market futures on spot market volatility for selected stocks from key industry sectors. The GARCH technique was used to capture the time varying nature of volatility of the Indian stock market. Tripathy & Rahman (2013) also use the GARCH model for forecasting daily stock volatility. Vegendla & Enke (2013) investigate the forecasting ability of Feedback Forward Neural Network using back propagation learning Recurrent Neural Networks and also GARCH models of historic volatility, implied volatility and model based volatility. The exercise is done for NASDAQ, DJIA, NYSE and S & P 500.

Panda & Deo (2014), Srinivasan & Prakasham (2014) and Srinivasan (2015), using different sets of variables, apply the

GARCH model or the Autoregressive Distributed Lag model, to understand the volatility spillover between various financial assets.

In a recent contribution, Dixit, Roy & Uppal (2013) have provided a framework for predicting India VIX using Artificial Neural Network. In their model, the first seven indicators are current day's open (CO), high (CH), low (CL) and close (CC) index values followed by previous day's high (PH), low (PL) and close (PC) index values. Next four input parameters were calculated using the simple moving average of the last (including the current day) 3 days (SMA3), 5 days (SMA5), 10 days (SMA10) and 15 days (SMA15) closing India VIX values.

McMillan (2004) presents a non-parametric framework for predicting implied volatility where Implied Volatility (IV) and Historic Volatility (HV) are grouped in deciles. This is discussed in detail in DattaChaudhuri & Sheth (2014) where such decilesare constructed for India VIX (IV) and standard deviation of NIFTY returns (HV). The methodology involves taking a 20 Day Moving Average (MA) of IV and a 10 Day, 20 Day and 50 Day Moving Averages of HV up to a date and constructing deciles. Then the actual values of the variables are computed on a subsequent date, outside the cut-off date, and the decile position of the values is marked off. This information is then used to execute options trading strategies and the results are discussed.

## The variables

Together with our methodology, our chapter differs from the existing literature in the choice of inputs and outputs. We do not take lagged values of volatility as the inputs. Further, we allow for two outputs namely volatility of NIFTY returns (NIFTYSDR) and volatility of gold returns (GOLDSDR). To calculate NIFTYSDR, we take 20 day rolling standard deviation, annualized, of NIFTY returns. This is historic

volatility and this is one of the variables that we want to predict. The other output is GOLDSDR which is also calculated as 20 day rolling standard deviation of gold returns, annualized.

As inputs we consider India VIX, CBOE VIX, volatility of crude oil returns (CRUDESDR), volatility of DJIA returns (DJIASDR), volatility of DAX returns (DAXSDR), volatility of Hang Seng returns (HANGSDR) and volatility of Nikkei returns (NIKKEISDR). As discussed earlier, INDIA VIX, as derived from the options market, is a forward looking indicator for actual volatility. So it finds place in our analysis as a predictor or input. We do not explicitly consider lagged values of NIFTYSDR as inputs, as the ANN framework would consider feedback from past values. Further, it would also allow for learning from future values. To allow for external shocks, as India is a large importer of crude oil, we consider CRUDESDR an input. In the recent past, political instability in the Middle East and related regions have impacted the expected availability of oil and has resulted in stock market instability in India. Global macroeconomic impacts have been incorporated through DJIASDR, DAXSDR, HANGSDR and NIKKEISDR. We have considered the impact of instability in both the western world and the eastern world. The inclusion of CBOE VIX is to factor in the impact of expected future volatility in the US market on the Indian market. That is, if CBOE VIX rises, some future instability in the US markets is foreseen. This in turn affects FII fund flows and hence NIFTYSDR.

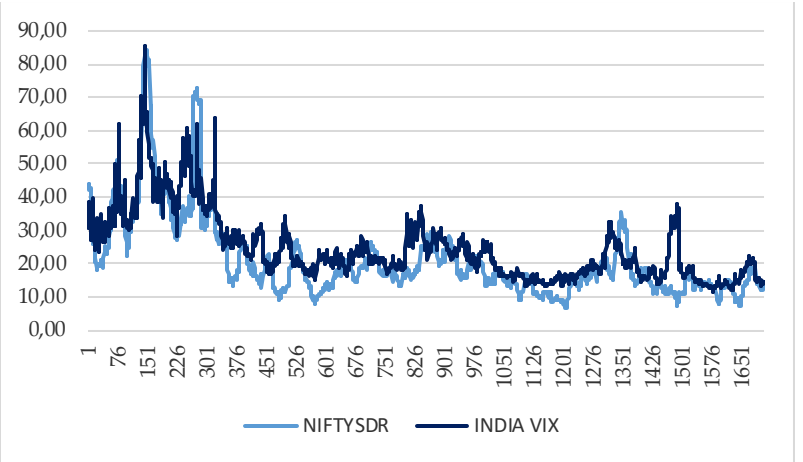


**Figure 2.** *INDIAVIX and NIFTYSDR for the period 3.3.2008 to 10.4.2015*

Figure 2clearly suggests that, overall, over a fairly long period, historic volatility and implied volatility do move together. So considering INDIA VIX as a predictor of NIFTYSDR is alright.

The following Figures 3 and 4show the movement in the two variables in different sub periods.

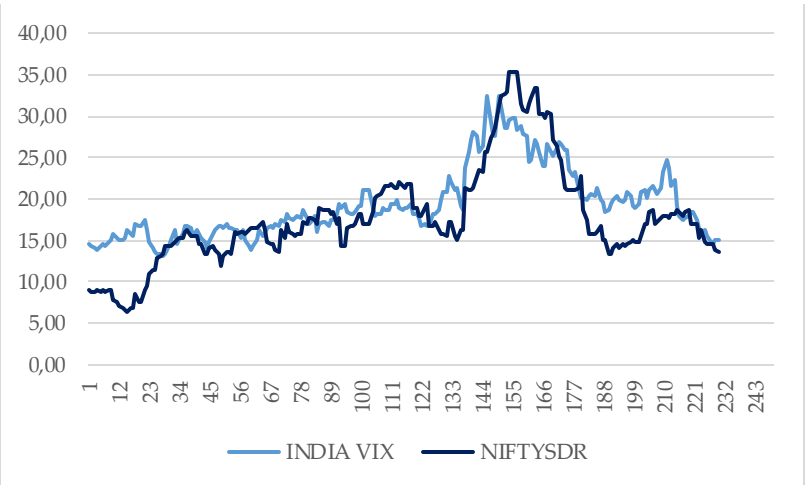


**Figure 3.** *INDIAVIX and NIFTYSDR for 2013*

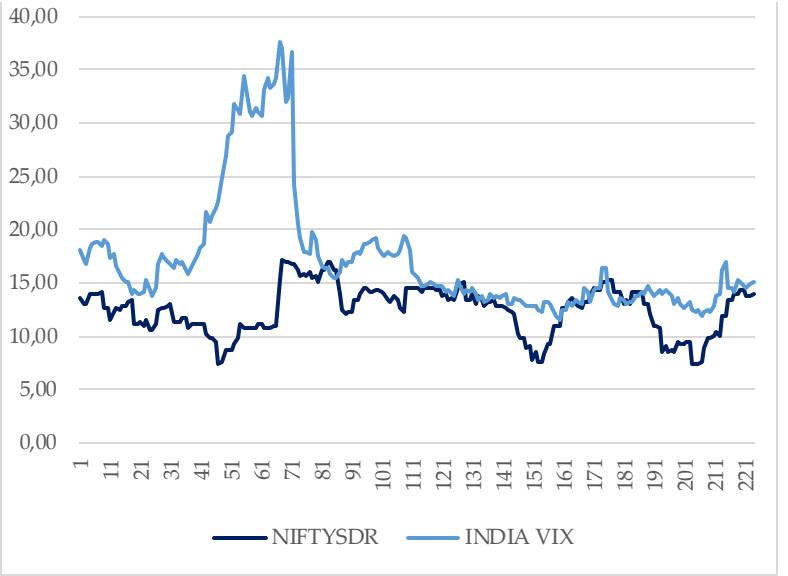


**Figure 4.** *INDIAVIX and NIFTYSDR for 2014*

Figures 3 and 4 reveal that for shorter time periods, the movements in the two variables are not always in tandem and hence the rationale for inclusion of other inputs in the analysis.

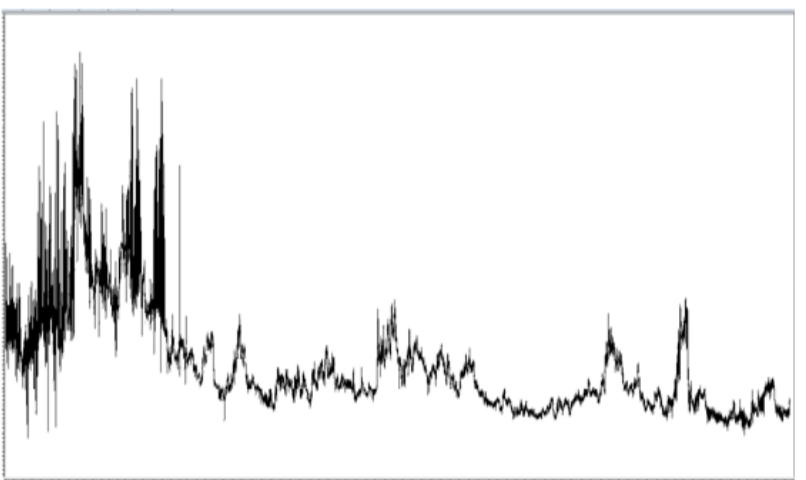

**Figure 5.** *India VIX for the period 5.3.2008 – 21.4.2015*

Figure 5 depicts the impact of global financial crisis of 2008 onINIDA VIX and clearly there are global factors that enter domestic expectations formation. This becomes even clear from Figure 6 where expected volatility in the US seems to go hand in hand with expected volatility in India. That is, global uncertainties affect US implied volatility, which in turn affects implied volatility index in India. There are, however, discrepancies, and hence both enter as inputs in our study.

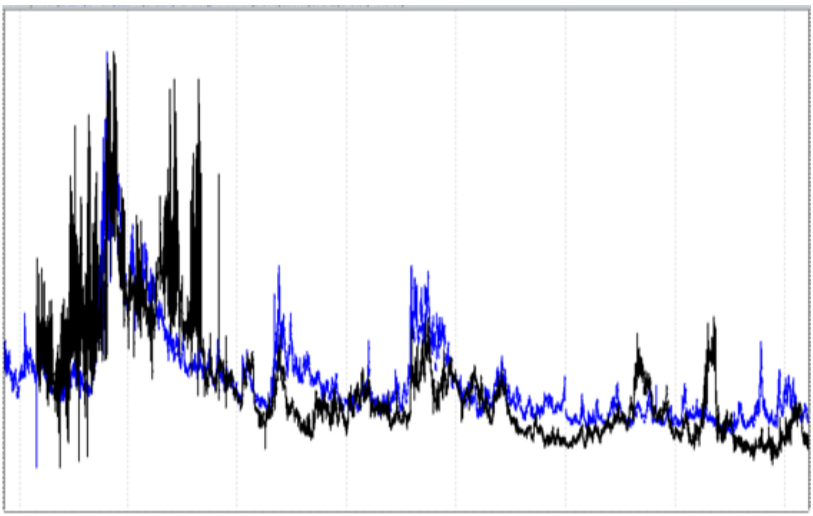

**Figure 6.** *INDIA VIX and CBOE VIX 2008 onwards*

## Results and analysis

In this chapter, three experiments have been conducted. In the first experiment, attempt has been made to forecast NIFTY Returns and Gold Returns for first four months of 2015 utilizing the entire daily data on the variables for the years 2013 and 2014 as training data. In experiment two, data for the entire year 2013 and a major part of 2014 has been used as training data to predict NIFTY Returns and Gold Returns for a part of 2014. In the third experiment, the

training data for the first experiment has been used to estimate NIFTY Returns and Gold Returns for a past period, year 2008. The latter has been done to examine whether the adopted framework can estimate market volatility of some past period based on the present scenario. This is our way of understanding the nature of the data and also the analytical framework used. It is also a means of validating our approach.

As discussed earlier, here, two different neural architecture and nine learning algorithms have been adopted. Number of hidden layers and their corresponding hidden neurons can be varied as well. For our experiments, only one hidden layer is used while number of hidden neurons has been varied at three levels (20, 30 & 40 respectively). Hence total number of trials is fifty four (2*9*3). Descriptive statistics of different performance indicators are presented to judge the results critically. Other important specifications of parameters which are used throughoutthe ANN modeling process have been shown in Table 1.

**Table 1.** *Important specification of parameters*

| Sl. No. | Parameter | Data/Technique Used |
|---|---|---|
| 1. | Number of input neuron(s) | Seven |
| 2. | Number of output neuron(s) | Two |
| 3. | Transfer function(s) | Tan-sigmoid transfer function (tansig) in hidden layer &purelin in output layer. |
| 4. | Error function(s) | Mean squared error(MSE) function |
| 5. | Type of Learning rule | Supervised learning rule |

**Source:** Authors' own construction

To evaluate the performance of the framework, three metrics namely mean squared error (MSE), regression coefficient (R) and mean absolute percentage error (MAPE) have been used. MSE is expressed as

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\{Y_{act}(i) - Y_{pred}(i)\}2$$

Regression coefficient (R) is the correlation measure between the actual and predicted outcomes which is computed as

$$R = \frac{N\left(\sum_{i=1}^{N} Y_{act}(i) * Y_{pred}(i)\right) - \left(\sum_{i=1}^{N} Y_{act}(i)\right) * \left(\sum_{i=1}^{N} Y_{pred}(i)\right)}{\sqrt{\left[N\sum_{i=1}^{N} Y_{act}(i)^2 - \left(\sum_{i=1}^{N} Y_{act}(i)\right)^2\right]\left[N\sum_{i=1}^{N} Y_{pred}(i)^2 - \left(\sum_{i=1}^{N} Y_{pred}(i)\right)^2\right]}}$$

MAPE is the average sum of absolute percentage error(s) over the entire dataset. Mathematically it is calculated as:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{Y_{act}(i) - Y_{pred}(i)}{Y_{act}(i)}\right| \times 100\%$$

where N denotes the total number of observations.

Descriptive statistics of all three performance indicators for experiment 1are shown in Tables 2 to 7.

**Table 2.** *Statistics of MSE of total 54 trials for the training dataset*

| MSE | Multi-Layer Feed Forward Network | Cascade Feed Forward Network |
|---|---|---|
| Min | 1.0528 | 1.8691 |
| Max | 3.9826 | 4.0125 |
| Average | 2.4627 | 2.6853 |
| Standard Deviation | 1.0424 | 1.2507 |

**Table 3.** *Statistics of R of total 54 trials for the training datasets*

| R | Multi-Layer Feed Forward Network | Cascade Feed Forward Network |
|---|---|---|
| Min | 0.9427 | 0.9358 |
| Max | 0.9842 | 0.9742 |
| Average | 0.9643 | 0.9543 |
| Standard Deviation | 0.0142 | 0.0205 |

**Table 4.** *Statistics of MAPEof total 54 trials for the training datasets*

| MAPE | Multi-Layer Feed Forward Network | Cascade Feed Forward Network |
|---|---|---|
| Min | 0.7653 | 0.8003 |
| Max | 1.0592 | 1.1203 |
| Average | 0.9021 | 0.9207 |
| Standard Deviation | 0.1071 | 0.1103 |

**Table 5.** *Statistics of MSEof total 54 trials for the testing datasets*

| MSE | Multi-Layer Feed Forward Network | Cascade Feed Forward Network |
|---|---|---|
| Min | 3.0244 | 3.5648 |
| Max | 4.1282 | 4.3206 |
| Average | 3.7251 | 3.9473 |
| Standard Deviation | 0.3581 | 0.3948 |

**Table 6.** *Statistics of R of total 54 trials for the testing datasets*

| R | Multi-Layer Feed Forward Network | Cascade Feed Forward Network |
|---|---|---|
| Min | 0.9534 | 0.9431 |
| Max | 0.9748 | 0.9763 |
| Average | 0.9654 | 0.9647 |
| Standard Deviation | 0.0068 | 0.0093 |

**Table 7.** *Statistics of MAPEof total 54 trials for the testing datasets*

| MAPE | Multi-Layer Feed Forward Network | Cascade Feed Forward Network |
|---|---|---|
| Min | 0.8328 | 0.8773 |
| Max | 1.1726 | 1.2016 |
| Average | 0.9592 | 1.018 |
| Standard Deviation | 0.1227 | 0.1904 |

MSE and MAPE values must be as low as possible toindicate efficient prediction; ideally a value of zero signifies no error. On the other hand, a value of R close to 1 is a must for strong prediction. For both training and test dataset, the values of performance indicatorsshown above justify the effectiveness of MLFF and CFFN tool as a forecasting tool for the problem at hand. It can thus be

concluded that volatility of NIFTY Returns and Gold Returns can be predicted using India VIX, CBOE VIX, CRUDESDR, DJIASDR, DAXSDR, HANGSDR and NIKKEISDR.

Figure 7 depicts the regression plot of forecast values as generated using the test data as against the actual data of 2015. The results indicate that the methodology used and the inputs chosen forecast the volatility of the outputs well.

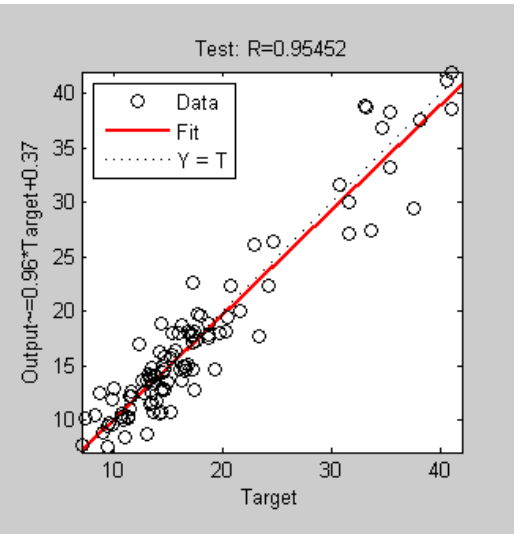


**Figure 7.** *Regression plot of Experiment 1.*
**Source:** MATLAB

For Experiment 2, we have kept the same experimental settings and the results are displayed in the following tables.

**Table 8.** *Statistics of MSE of total 54 trials for the training datasets*

| MSE | Multi-Layer Feed Forward Network | Cascade Feed Forward Network |
|---|---|---|
| Min | 2.8214 | 2.9403 |
| Max | 4.3251 | 3.8923 |
| Average | 3.6284 | 3.5981 |
| Standard Deviation | 0.5065 | 0.4818 |

**Table 9.** *Statistics of R of total 54 trials for the training datasets*

| R | Multi-Layer Feed Forward Network | Cascade Feed Forward Network |
|---|---|---|
| Min | 0.9432 | 0.9581 |
| Max | 0.9872 | 0.9868 |
| Average | 0.9604 | 0.9677 |
| Standard Deviatior | 0.0158 | 0.0139 |

**Table 10.** *Statistics of MAPEof total 54 trials for the training datasets*

| MAPE | Multi-Layer Feed Forward Network | Cascade Feed Forward Network |
|---|---|---|
| Min | 0.8603 | 0.8018 |
| Max | 1.1209 | 1.0962 |
| Average | 1.0062 | 0.9623 |
| Standard Deviation | 0.0935 | 0.0968 |

**Table 11:** *Statistics of MSEof total 54 trials for the testing datasets*

| MSE | Multi-Layer Feed Forward Network | Cascade Feed Forward Network |
|---|---|---|
| Min | 3.1267 | 3.2207 |
| Max | 4.2582 | 4.1263 |
| Average | 3.7508 | 3.5262 |
| Standard Deviation | 0.4236 | 0.3708 |

**Table 12.** *Statistics of R of total 54 trials for the testing datasets*

| R | Multi-Layer Feed Forward Network | Cascade Feed Forward Network |
|---|---|---|
| Min | 0.9268 | 0.9325 |
| Max | 0.9634 | 0.9662 |
| Average | 0.9483 | 0.9518 |
| Standard Deviation | 0.0143 | 0.0127 |

**Table 13.** *Statistics of MAPEof total 54 trials for the testing datasets*

| MAPE | Multi-Layer Feed Forward Network | Cascade Feed Forward Network |
|---|---|---|
| Min | 0.8457 | 0.8223 |
| Max | 1.1063 | 1.0218 |
| Average | 0.9641 | 0.9414 |
| Standard Deviation | 0.0906 | 0.0892 |

Figure 8 again indicates that for the second experiment also the methodology used and the inputs chosen forecast the volatility of the outputs well.

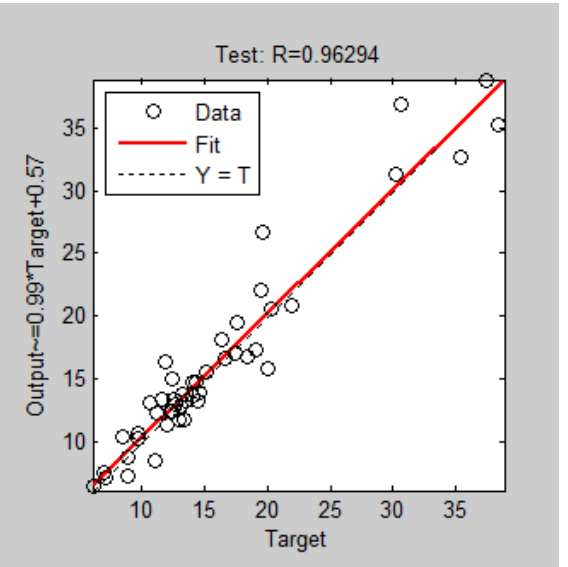


**Figure 8.** *Regression plot of Experiment 2*
**Source:** MATLAB

The third experiment that we perform is interesting as we have been employed the data for 2013 and 2014 together as training data to estimate the volatility back in 2008. Tables 14-19 portray the results and the findings are discussed later.

**Table 14.** *Statistics of MSE of total 54 trials for the training datasets*

| MSE | Multi-Layer Feed Forward Network | Cascade Feed Forward Network |
|---|---|---|
| Min | 14.2362 | 15.6243 |
| Max | 22.3898 | 23.1684 |
| Average | 19.7424 | 20.1871 |
| Standard Deviation | 2.3516 | 2.2783 |

**Table 15.** *Statistics of R of total 54 trials for the training datasets*

| R | Multi-Layer Feed Forward Network | Cascade Feed Forward Network |
|---|---|---|
| Min | 0.8891 | 0.8934 |
| Max | 0.9251 | 0.9362 |
| Average | 0.9054 | 0.9126 |
| Standard Deviation | 0.0107 | 0.0125 |

**Table 16.** *Statistics of MAPE of total 54 trials for the training datasets*

| MAPE | Multi-Layer Feed Forward Network | Cascade Feed Forward Network |
|---|---|---|
| Min | 5.7682 | 5.5803 |
| Max | 8.0974 | 7.9561 |
| Average | 6.6785 | 6.4327 |
| Standard Deviation | | |

**Table 17.** *Statistics of MSE of total 54 trials for the testing datasets*

| MSE | Multi-Layer Feed Forward Network | Cascade Feed Forward Network |
|---|---|---|
| Min | 16.2218 | 15.9583 |
| Max | 22.3735 | 22.5612 |
| Average | 18.8187 | 19.0462 |
| Standard Deviation | 2.0846 | 2.1283 |

**Table 18.** *Statistics of R of total 54 trials for the testing datasets*

| r | Multi-Layer Feed Forward Network | Cascade Feed Forward Network |
|---|---|---|
| Min | 0.8772 | 0.8806 |
| Max | 0.9184 | 0.9189 |
| Average | 0.8923 | 0.8918 |
| Standard Deviation | 0.0138 | 0.0126 |

**Table 19.** *Statistics of MAPE of total 54 trials for the testing datasets*

| MAPE | Multi-Layer Feed Forward Network | Cascade Feed Forward Network |
|---|---|---|
| Min | 6.0182 | 6.0143 |
| Max | 8.3264 | 8.7065 |
| Average | 7.2165 | 7.5781 |
| Standard Deviation | 0.77 | 0.7981 |

Interestingly in this experiment, it can be seen that the average MSEand MAPE values for both training and testing set are considerably larger in compared to the earlier experiments. Similarly average R values are also lower in both training and testing set. So it may be inferred that prediction accuracy of the model trained in present time has goes down when asked to forecast market volatility back in 2008. Given the extent of enormously increased volatility in 2008 post the crisis, as shown in Figure 5, the results are not very surprising. The regression plot in Figure 9 also captures this.

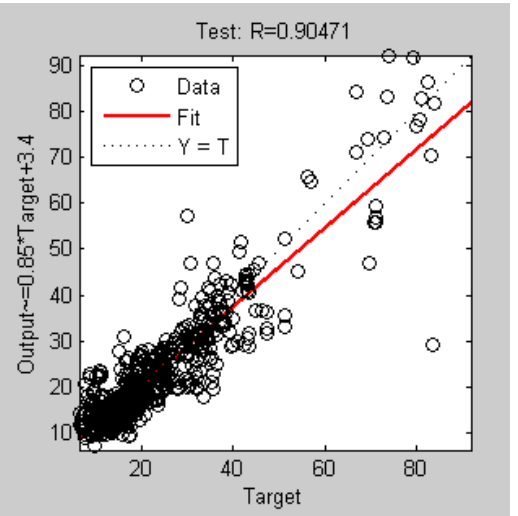


**Figure 9.** *R.*
**Source:** *MATLAB*

In Figures 10, 11 and 12, we portray the performance of the two different neural architectures on test data set. A comparative analysis of MLFF and CFFN has also been carried out to statistically analyze their performance. Statistical t-test has been conducted on MSE to judge whether their performances are significantly different or not. Table 20 depicts the outcomes.

**Table 20.** *Significance values (on test cases)*

| Experiment 1 | Experiment 2 | Experiment 3 |
|---|---|---|
| 0.221 (two tailed) | 0.305 (two tailed) | 0.184 (two tailed) |

**Source:** Authors' own construction

As none of the values of the test statistic are significant, it can be concluded that there is no significant difference in performance among two models.

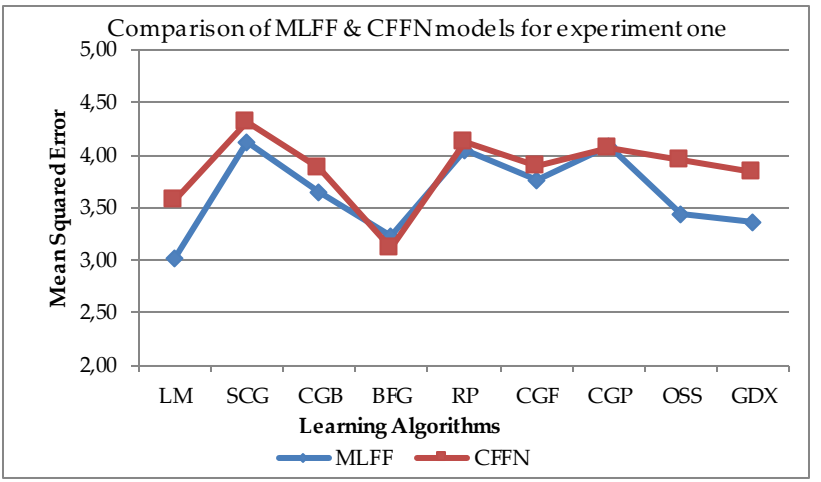


**Figure 10.** *EXPERIMENT 1*

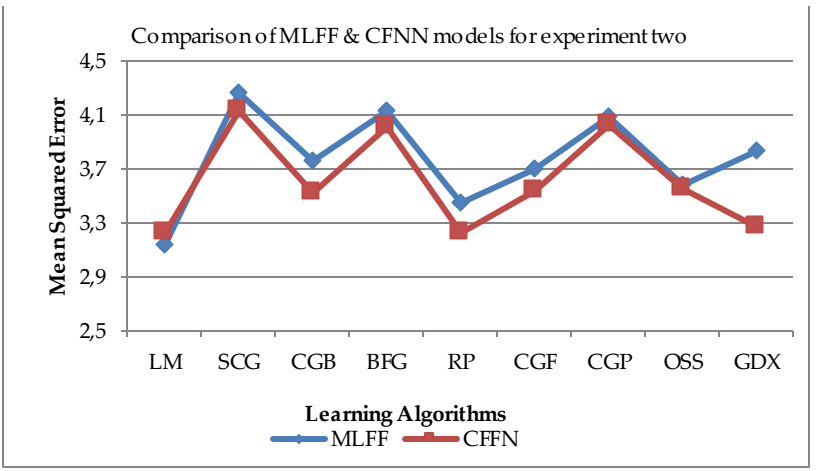


**Figure 11.** *EXPERIMENT 2*

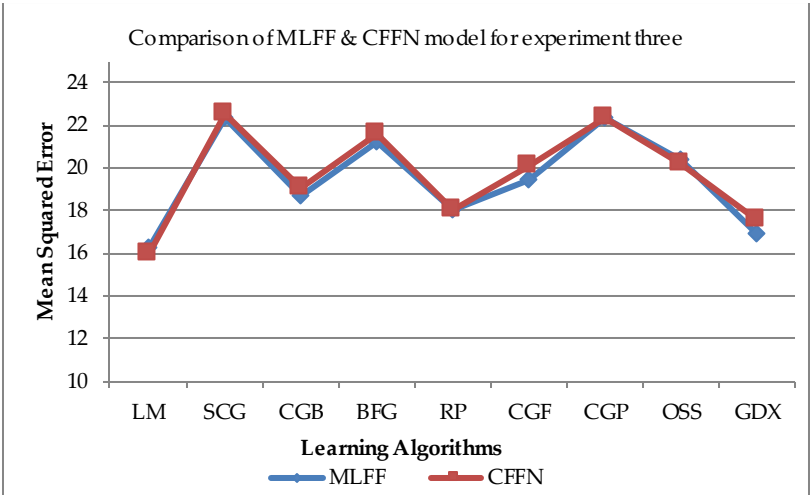


**Figure 12.** *EXPERIMENT 3*

## Conclusion

The purpose of this chapter was to examine the efficacy of the ANN framework in predicting volatility in the Indian stock market. We used a multiple input multipleoutput structure using two different neural architecture and nine learning algorithms. For our experiments, only one hidden layer was used while number of hidden neurons has been varied at three levels (20, 30 & 40 respectively). Hence total number of trials was fifty four (2*9*3). We conducted our exercise for three different time periods. Our framework could satisfactorily forecast volatility for 2015 using training data for 2013-14. However, the prediction accuracy of the model, trained in present time, has goes down when asked to forecast market volatility back in 2008.

## References

Adhikari, R. (2015). A neural network based linear ensemble framework for time series forecasting, *Neurocomputing*, 157, 231-242. doi. 10.1016/j.neucom.2015.01.012

Aish, A.M., Zaqoot H.A., & Abdeljawad, S.M. (2015). Artificial neural network approach for predicting reverse osmosis desalination plants performance in the Gaza Strip, *Desalination*, 367, 240-247. doi. 10.1016/j.desal.2015.04.008

Datta Chaudhuri, T., & Kinjal, S. (2014). Forecasting volatility, volatility trading and decomposition by Greeks, *CBS Journal of Management Practices*, 1(1), 59-70.

Dixit, G., Roy, D., & Uppal, N. (2013). Predicting India volatility index: An application of artificial neural, *Network*, 70(4), 22-30. doi. 10.5120/11950-7768

Ghiassi, M., Lio, D., & Moon, B. (2015). Pre-production forecasting of movie revenues with a dynamic artificial neural network, *Expert Systems with Applications*, 42(6), 3176-3193. doi. 10.1016/j.eswa.2014.11.022

Karnik, S.R., Gaitonde, V.N., Campos-Rubio, J., Esteves- Correia, A., Abrão, A.M., & Paulo-Davim, J. (2008). Delamination analysis in high speed drilling of carbon fiber reinforced plastics (CFRP) using artificial neural network model, *Materials and Design*, 29(9), 1768–1776. doi. 10.1016/j.matdes.2008.03.014

Lasheras, F.S., Juez, de cosJuez, F.J., Sanchez, A.S., Krzemie, A., & Fernandez, P.R. (2015). Forecasting the COMEX copper spot price by means of neural networks and ARIMA models, *Research Policy*, 45, 37-43. doi. 10.1016/j.resourpol.2015.03.004

Malhotra, G. (2012). Impact of futures trading on volatility of CNX nifty, *IIMS Journal of Management Science*, 3(1), 166-178.

Malliaris, M., & Salchenberger, L. (1996). Using neural networks to forecast the S&P 100 implied volatility, *Neurocomputing*, 10(2), 183-195. doi. 10.1016/0925-2312(95)00019-4

McMillan, L.G. (2004). *McMillan on Options*, John Wiley & Sons, Inc., Hoboken, New Jersey.

Ndaliman, M.B., Hazza, M., Khan, A.A., & Ali. M.Y. (2012). Development of a new model for predicting EDM properties of Cu-TaC compact electrodes based on artificial neural network

method, *Australian Journal of Basic and Applied Sciences*, 6(13), 192-199.

Oko, E., Meihong, W., & Zhang, J. (2015). Neural network approach for predicting drum pressure and level in coal-fired subcritical power plant, *Fuel*, 151, 139-145. doi. 10.1016/j.fuel.2015.01.091

Pal, M., Pal, S.K., & Samantaray, A.K. (2008). Artificial neural network modeling of weld joint strength prediction of a pulsed metal inert gas welding process using arc signals. *Journal of Materials Processing Technology*, 202(1-3), 464–474. doi. 10.1016/j.jmatprotec.2007.09.039

Panda, P., & Deo, M. (2014). Asymmetric and volatility spillover between stock market and foreign exchange market: Indian experience, *IUP Journal of Applied Finance*, 20(4), 69-82.

Passarelli, D. (2008). *Trading Options Greeks*, Bloomberg Press, New York.

Ramasamy, P., Chandel, S.S., & Yadav, A.K. (2015). Wind speed prediction in the mountainous region of India using an artificial neural network model, *Renewable Energy*, 80, 338-347. doi. 10.1016/j.renene.2015.02.034

Rather, A.M., Agarwal, A., & Sastry, V.N. (2015). Recurrent neural network and a hybrid model for prediction of stock returns, *Expert Systems with Applications*, 42(6), 3234-3241. doi. 10.1016/j.eswa.2014.12.003

Srinivasan, P., & Karthigen, P. (2014). Gold price, stock price and exchange rate nexus: The case of India, *IUP Journal of Financial Risk Management*, 11, 52-62.

Srinivasan, P. (2015). Modelling and forecasting time-varying conditional volatility of Indian stock market, *IUP Journal of Financial Risk management*, 12(3), 49-64. doi. 10.1515/saeb-2017-0021

Tripathy, S., & Rahman, A. (2013). Forecasting daily stock volatility using garch model: A comparison between BSE and NSE, *IUP Journal of Applied Finance*, 15, 71-83.

Vejendla, A., & Enke, D. (2013). Evaluation of GARCH, RNN & FNN models for forecasting volatility in the financial markets, *IUP Journal of Financial Risk Management*, 10(1), 41-49.

Walczak, S., & Sincich, T. (1999). A comparative analysis of regression and neural networks for university admissions. *Information Sciences*, 119(1-2), 1-20. doi. 10.1016/S0020-0255(99)00057-2

Zhao, Y., Du, X., Xia, G., & Wu., L. (2015). A novel algorithm for wavelet neural network s with application to enhanced PID controller designs, *Neurocomputing*, 158, 257-267. doi. 10.1016/j.neucom.2015.01.015

**For Cited this Chapter:**

**Chaudhuri, T.D., & Ghosh, I. (2020). Forecasting volatility in Indian stock market using artificial neural network with multiple inputs and outputs, *In* T.D. Chaudhuri (Ed.), *Overview of the Indian Economy: Empirical analysis of financial markets, exchange rate movements and industrial development* (pp.113-140), KSP Books: Istanbul.**

ISBN: 978-605-7736-93-2 (e-Book)
KSP Books 2020
© KSP Books 2020

**Copyrights**

Copyright for this Book is retained by the author(s), with first publication rights granted to the Book. This is an open-access Book distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by-nc/4.0 ).

# 6 Artificial neural network and time series modeling based approach to forecasting the exchange rate in a multivariate framework

*By*
Tamal DATTA CHAUDHURI
& Indranil GHOSH

## Introduction

Exchange rate is the price of foreign currency in terms of the domestic currency. If 1 US Dollar can buy 65 Indian Rupees, then the exchange rate is either Rs.65/$1 or $1/Rs.65. As exchange rate is a price, any analysis of it has to be based on factors affecting demand and supply of foreign currency. Foreign currency flows in and out of an economy through both the current account and the capital account of the balance of payments [for a lucid explanation see Caves & Jones (1973)]. Export and import of goods and services leads to inflow and outflow respectively of foreign currency through the current account. Portfolio flows and foreign direct investment lead to inflow or outflow through the capital account. Since balance of payments have to balance, any discrepancy is taken care of by change in the level of foreign currency reserves.

In today's world, most economies have shifted to a flexible exchange rate system where the exchange rate adjusts to clear the market. In case of abnormal movements in the rate, central banks do intervene to stabilize the currency by selling from reserves or buying from the market. Since the exchange rate is not entirely in the hands of the domestic economy, and since not all foreign currency contracts are spot contracts, exporters and importers face the risk of adverse movements in the exchange rate in the future. For this, there is a strong forward and futures market in the foreign currency market. All the above characteristics of this market has led to the emergence of three sets of players namely, agents with underlying interest in foreign currency like exporters, importers and traders, speculators and arbitrageurs.

Exports and imports of an economy depend on the economic background and the various policies for economic development, the technological capabilities, natural endowment of resources, and the sociological and demographic features. India does not have large reserves of crude oil. Hence around 80% of its imports is crude oil. Venezuela on the other hand has significant reserves of crude oil and hence it exports mostly oil. Brazil is endowed with coffee plantations and have nurtured them to make it the largest coffee producer and exporter in the world. So movement in foreign currency through the current account are dependent on factors that affect exports and imports. Movement of foreign exchange through the capital account has to do with returns on foreign exchange, and hence on the relative attractiveness of two nations in terms of returns.

Let there be two countries USA (u) and India (i), the rupee dollar spot rate be $S_p$, the rupee dollar forward rate be $S_f$ and the nominal rates of interest in the two countries be $R_u$ and $R_i$ respectively. Then, in equilibrium, the following equality has to hold.

$$(1 + R_i) = S_f(1 + R_a)/S_p$$

Otherwise funds will move between countries. This condition is known as the "covered interest arbitrage" condition. The condition states that, Rs.1 in India, converted at the spot rate to US dollars, invested in the US for a period and brought back to India by the period forward rate today, has to be equal to the returns from the Indian market for the period. If the LHS is greater than the RHS, then foreign funds will move into India and if RHS is greater than LHS then funds will move from India to the US. If we replace nominal rates of interest with real rates of interest and as real rates will equalize across countries, the funds movement will be governed by the rates of inflation in the two countries.

We mentioned that the returns from the two countries have to be equal in equilibrium, but this equality has to be qualified by the ratio of spot rate to the forward rate. This is the characteristic of the exchange rate market which attracts arbitrageurs. Notice that the above equality condition is one equation in four unknowns. Thus given the spot rate and the rates of return in the two countries, the forward rate can be derived. On the other hand, if we use the quoted futures rate, then one of the interest rates can be derived. This derived interest rate is called the Implied Repo Rate, and if it is different from the actual interest rate, will give rise to arbitrage.

The above discussion was intended to establish that any study on exchange rate movements and forecasting, has to include explanatory variables from both the current account and the capital account. In this chapter, we include such factors to forecast the value of the Indian rupee vis a vis the US Dollar. There could be certain other factors like political instability and lack of mechanism for enforcement of contracts that can affect both direct foreign investment and

also portfolio investment. In this chapter we include such variables also.

To forecast the exchange rate, we have used two different classes of frameworks namely, Artificial Neural Network (ANN) based models and Time Series Econometric models. Multilayer Feed Forward Neural Network (MLFFNN) and Nonlinear Autoregressive models with Exogenous Input (NARX) Neural Network are the approaches that we have used as ANN models. Generalized Autoregressive Conditional Heteroskedastic (GARCH) and Exponential Generalized Autoregressive Conditional Heteroskedastic (EGARCH) techniques are the ones that we use as Time Series Econometric methods.

Although the existing literature on forecasting the exchange rate is quite rich and this chapter is embedded in that literature, the following features of the chapter will add to the understanding of the problem at hand and also refine the forecasting of future exchange rate movements. First, we use daily data of the exchange rate and the other explanatory variables. We do not incorporate any macroeconomic variables. Our contention is that the exchange rate is a price of a financial asset which is traded on real time basis. Hence its forecast should use such variables, the data on which are also available daily. Second, we use both current account and capital account factors as explanatory variables. In particular, we use the forward rate as an explanatory variable. Third, as mentioned above, we incorporate explanatory variables that reflect the economic/political/ financial instability of an economy. Fourth, in contrast to many papers that have used machine learning techniques and econometric techniques in a univariate framework, ours is a multivariate approach. Fifth, we apply both traditional econometric and machine learning tools in a multivariate framework, which enables us to compare the efficiency of

these two classes of models. Sixth, application of the NARX model is quite unique.

The plan of the chapter is as follows. A brief literature survey is presented in Section 2. Detailed description of the dataset and methodologies are elucidated in Section 3. Results obtained from ANN modelling and time series modelling are explained in Sections 4 and 5 respectively. Comparative analysis of the performance of the two different frameworks ispresented in Section 6. Section 7 concludes the chapter.

## Literature review

Meese & Rogoff (1983) presented a model of forecasting the exchange rate which incorporated the characteristics of the flexible price monetary model (Frenkel-Bilson), the sticky price monetary model (Dornbusch-Frankel), and the sticky price asset model (Hooper – Morton). The relationship they postulated was

$$S = A + A1\ (M - M^*) + A2\ (Y - Y^*) + A3(Rs - Rs^*) + A4\ (P - P^*) + A5\ TB + A6\ TB^* + u$$

where S is log dollar price of foreign currency, M - M* is log ratio of domestic and foreign money supply, Y – Y* is log ratio of domestic and foreign income, Rs – Rs* and P – P* are the short term interest rate and inflation rate differential respectively, TB and TB* are the foreign exchange reserves and u is the random disturbance term. Their specification is macroeconomic in nature where money supply, national income and forex reserves were considered.

Zhang & Berardi (2001) used neural network ensembles for predicting theexchange rate between the British pound and US dollar. In this chapter, the inputs are the lag variables of the output and the focus is on the technique of prediction.

Perwej & Perwej (2012) considered forecasting the Indian Rs/UD$ exchange rate using the Artificial Neural Network framework. Here also the focus is on selection of number of input nodes and hidden layers. Lagged values of the exchange rate is used to predict the future values of the exchange rate.

In the lines of Meese & Rogoff (1983), Lam, Fung & Yu (2008) consider a Bayesian model for forecasting the exchange rate. The explanatory variables they consider are quite exhaustive and include stock price, change in stock price, long-term interest rate, short-term interest rate, term spread, oil price, change in oil price, exchange rate return of the previous period, sign of exchange rate return of the previous period, seasonally adjusted real GDP, change in seasonally adjusted real GDP, seasonally adjusted money supply, change in seasonally adjusted money supply, consumer price level, inflation rate, and ratio of current account to GDP. Many of the variables are measured relative to that of the foreign country. However, the forward rate is missing as an explanatory variable.

Ravi, Lal & Raj Kiran (2012) use sixnonlinear ensemble architectures for forecasting exchange rates. Although the paper applies many techniques for forecasting a number of exchange rates, the input variables are only the lagged values of the exchange rate itself.

Dua & Ranjan (2011) applied both Vector Auto Regression (VAR) and Bayesian Vector Auto Regression (BVAR)for forecasting the Indian Rupee/US Dollar exchange rate. Following Meese & Rogoff (1983), they also consider both current account and capital account variables along with forward exchange rates. While they find that the BVAR outperforms VAR, they also observe that forecast accuracy improves if we include the forward premium and volatility of capital flows.

Pacelli (2012) analyzed and compared the predictive ability of ANN, ARCH and GARCH models where the output variable was the daily exchange rate Euro/Dollar and the input variables were the Nasdaq Index, Daily Exchange Rate Eur/Usd New Zealand, Gold Spot Price USA, Average returns of Government Bonds - 5 years in the USA zone, Average returns of Government Bonds - 5 years in the Eurozone, Crude Oil Price, Exchange rate Euro / US dollar of the previous day compared to the day of the output.

Imam *et al.* (2015) presented a comprehensive survey work on various computational intelligent methods and financial quantitative models used in predictive modelling of exchange rates. The study highlights the usage of Artificial Neural Networks, Support Vector Machine, ARCH/GARCH models etc. in the particular area.

Androu & Zombanakis (2006) utilized Artificial Neural Network based approaches to forecast the Euro exchange rate versus the United States Dollar and the Japanese Yen. Their work suggested the presence of random behavior of time series according to Rescaled Range Statistic (R/S). However the NN model adopted in their study indeed resulted in good prediction in terms of Normalized Root Mean Squared Error (NRMSE), the Correlation Coefficient (CC), the Mean Relative Error (MRE), the Mean Absolute Error (MAE) and the Mean Square Error (MSE).

Vojinovic & Kecman (2001) employed Radial Basis Function Neural Network to predict daily $US/$NZ closing exchange rates. Findings indicated that Radial Basis Function Neural Network outperformed traditional autoregressive models.

Garg (2012) presented a novel framework comprising of GARCH extended machine learning models namely, Regression trees, Random Forests, Support Vector Regression (SVR), Least Absolute Shrinkage and Selection Operator (LASSO) and Bayesian Additive Regression trees

(BART) to predict EUR/SEK, EUR/USD and USD/SEK exchange rates on both monthly and daily basis in a multivariate framework where different sets of predictors were utilized for prediction of respective exchange rates.

Jena *et al.* (2015) used Knowledge Guided Artificial Neural Network (KGANN) structure for exchange rate prediction. Premanode & Toumazou (2013) proposed a novel methodology where differential Empirical Mode Decomposition (EMD) is utilized to improve the performance of Support Vector Regression (SVR) to forecast exchange rates. Findings suggested that the proposed framework outperformed Markov Switching GARCH and Markov Switching Regression Models. Majhi *et al.* (2012) made a comparative study using Wilcoxon Artificial Neural Network (WANN) and Wilcoxon Functional Link Artificial Neural Network (WFLANN) in predictive modelling of exchange rate.

There is an extant literature in finance where both machine learning tools and econometric methods have been applied. For example, Datta Chaudhuri & Ghosh (2015) applied multilayer feedforward neural network and cascaded feedforward neural network to predict volatility measured in terms of volatility in NIFTY returns and volatility in gold returns over the years. Predictors considered in the study were India VIX, CBOE VIX, volatility of crude oil returns, volatility of DJIA returns, volatility of DAX returns, volatility of Hang Seng returns and volatility of Nikkei returns. Results justified the usage of ANN based methodology.

Bhat & Nain (2014) used GARCH, EGARCH and Component GARCH (CGARCH) to critically analyze volatility measures of four different Bombay Stock Exchange (BSE) indices during January, 1, 2002 to December, 31, 2013. Findings reported that BSE IT, BSE PSU, BSE Metal and BSE Bankex exhibit clear volatility clustering.

Tripathy & Rahman (2013) used GARCH (1, 1) model to measure the conditional market volatility of Bombay Stock Exchange (BSE) and Shanghai Stock Exchange (SSE). Empirical results obtained from daily data of 23 years of their study strongly indicate presence of volatility in the market.

## Data and methodology

The data set chosen for the study is daily data from 1.1.2009 to 8.4.2016 with 1783 observations. The dependent variable is the Rupee Dollar exchange rate (FX1). The independent variables are the 3 monthRupee Dollar futures exchange rate (FX4), NIFTY returns (NIFTYR), Dow Jones Industrial Average returns (DJIAR), Hang Seng returns (HSR), DAX returns (DR), crude oil price (COP), CBOE VIX (CV) and India VIX (IV). Inclusion of FX4, NIFTYR, DJIAR, HSR and DR as explanatory variables follows from the "covered interest arbitrage condition". While NIFTYR represents daily returns from the Indian stock market, we have included DJIAR and DR to represent returns from the western part of the world and HSR to represent returns in the eastern part of the world. CV and IV have been included to control for relative uncertainty in the Indian market vis a vis the US market as they are measures of Implied Volatility and are forward looking measures. As crude oil is the single largest import item of India, COP is the included to represent the current account in the balance of payments. In this study, thus, we have both current account and capital account variables, along with measures of volatility. Descriptive Statistics of all these variables are presented in the following table.

**Table 1.** *Descriptive Statistics of Variables*

| Variable | Mean | Median | Maximum | Minimum | Std. Dev. | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|
| FX1 | 54.6283 | 54.2300 | 69.0625 | 43.9150 | 7.5670 | 0.1598 | 1.5823 |
| FX4 | 55.0946 | 55.0588 | 69.8575 | 44.3800 | 8.0879 | 0.1485 | 1.4841 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| NIFTYR | 0.00059 | 0.00047 | 0.17744 | -0.01681 | 0.0128 | 1.3702 | 24.3389 |
| DJIAR | 0.00043 | 0.00053 | 0.06835 | -0.05546 | 0.01037 | -0.1089 | 7.0345 |
| HSR | 0.00029 | 0.00002 | 0.05756 | -0.06461 | 0.01364 | -0.11242 | 5.2778 |
| DR | 0.00051 | 0.00089 | 0.08152 | -0.05819 | 0.01417 | -0.00099 | 5.19045 |
| COP | 0.00039 | 0.00027 | 0.26874 | -0.09007 | 0.02187 | 1.69542 | 19.9603 |
| CV | 20.2721 | 17.5950 | 137.150 | 10.3200 | 8.43918 | 2.79998 | 24.8811 |
| IV | 22.1800 | 19.9450 | 63.5800 | 11.5650 | 8.08505 | 1.71638 | 6.28577 |

The values of skewness and kurtosis measures of majority of the variables indicate the presence of leptokurtosis. Jarque-Bera test has been conducted for statistical significance. The results shown in Table 2 dispel the normality assumption at 0.1% level for all the variables, indicating the presence of volatility, thus justifying the use of GARCH and EGARCH models.

**Table 2.** *Jarque-Bera Test*

| Variable | Jarque-Bera | p-value | Significance |
|---|---|---|---|
| FX1 | 156.8169 | 0.000000 | *** |
| FX4 | 177.1663 | 0.000000 | *** |
| NIFTYR | 34367.26 | 0.000000 | *** |
| DJIAR | 1212.111 | 0.000000 | *** |
| HSR | 388.9902 | 0.000000 | *** |
| DR | 356.2574 | 0.000000 | *** |
| COP | 22211.97 | 0.000000 | *** |
| CV | 37877.93 | 0.000000 | *** |
| IV | 1676.577 | 0.000000 | *** |

*** Significant at 1% level

## Multilayer Feed-Forward Network (MLFFNN)

It is a standard Artificial Neural Network (ANN) technique that attempts to mimic the working nature of the human brainto extract the hidden pattern between a set of inputs and outputs (Haykin, 1999). Human brain is composed of around $10^{10}$ number of highly interconnected units known as neurons. Similarly neurons are structured and connected in a hierarchical manner in a layered architecture of ANN. There are three distinct interconnected

layers in a typical ANN architecture namely, an input layer, hidden layer(s) and an output layer. They are connected via neurons and strength of each connection is actually represented by numeric weight value. For prediction tasks, basically these weight values corresponding to decision boundary, are estimated using various optimization algorithms on training data set. Once the estimated values are stabilized after validation, trained ANN is tested against a test data set to assess its predictive power.

A Multilayer Feed-Forward Neural Network (MLFFNN) is also composed of an input layer, one or more hidden layers and an output layer. A typical MLFFNN having 5 inputs is depicted below.

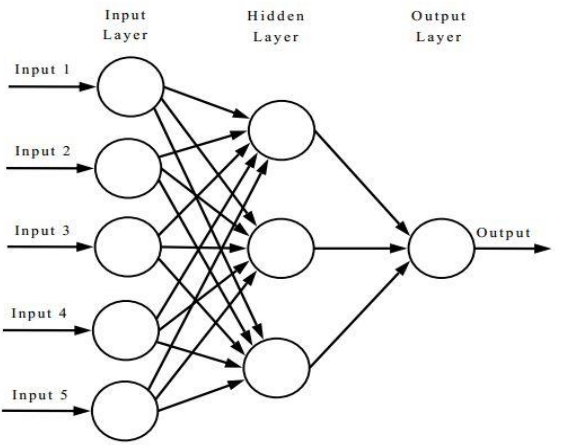


**Figure 1.** *Simple architecture of MLFFNN*

Input layer consists of simultaneously fed input units. Subsequently weighted inputs are fed into hidden layer. The weighted outputs of hidden layer(s) serve as the inputs to output layer and represent the prediction of network. As none of the weights cycle back to an input unit or to a previous layer's output unit and there are at least three

distinct layers, this network topology is called Multilayer Feed-Forward network.

Each output unit in a particular layer of MLFFNN considers weighted sum of outputs from previous layer's as inputs. It then applies a nonlinear function to the received input and forwards them to the subsequent layer. Each input signal ($x_i$) is associated with a weight ($w_i$). The overall input I to the processing unit is a function of all weighted inputs given by

$$I = f(\sum x_i \times w_i) \quad (1)$$

The activation state of the processing unit (A) at any time is a function (usually nonlinear) of I.

$$A = g(I) \quad (2)$$

In general, logistic or sigmoid function is used as activation function. If net input to unit j is $I_j$ then output ($O_j$) of unit j may be calculated as:

$$O_j = \frac{1}{1+e^{-I_j}} \quad (3)$$

The output Yfrom the processing unit is determined by the transfer function h

$$Y = h(A) = h(g(I)) = h(g(f(\sum x_i \times w_i)$$
$$= \Theta(\sum x_i \times w_i) \quad (4)$$

Given large sample of training data, MLFFNN can estimate the weight values and perform nonlinear regression. Once the estimated values are stabilized after validation, trained MLFFNcan be utilized for prediction on test data set. Backpropagation algorithm has been used as

training algorithm of MLFFNN for estimation of parameters to capture the nonlinear pattern between set of outputs and inputs for predictive modelling.

We now elucidate the working principle of Backpropagation Algorithm as reported by Han *et al*. (2009). Steps of general purpose backpropagation algorithm are outlined below.

1. Randomly initialize the weight and bias values of the network.
2. While terminating condition is not met {
3. For each training sample in dataset {
4. For each input layer unit j {
5. Output of an input unit is $O_j = I_j$; //
6. For each hidden or output layer unit j {
7. Net input of unit j is computed as $I_j = \sum_i w_{ij} O_i + \theta_j$; // $\theta_j$ is the respective bias value
8. Output of each unit j $O_j = \frac{1}{1+e^{-I_j}}$; }
9. For each unit j in the output layer {
10. Error is calculated as: $Err_j = O_j(1-O_j)(T_j-O_j)$; // $T_j$ is the target at $j^{th}$ unit }
11. For each unit j in the hidden layer(s), from last to first hidden layer {
12. Error is computed as: $Err_j = O_j(1-O_j)\sum_k Err_k w_{jk}$ ; }
13. For each weight $w_{ij}$ in network {
14. Weight value as are updated as: $w_{ij} = w_{ij} + \Delta w_{ij}$ ;
15. Where $\Delta w_{ij} = (l) Err_j O_i$ ; // l denotes the learning rate }
16. For each bias $\theta_j$ in network {
17. Bias values are modified as: $\theta_j = \theta_j + \Delta\theta_j$;
18. Where $\Delta\theta_j = (l) Err_j$ ; }
19. }}

## NARX (Nonlinear Autoregressive models with exogenous input) Neural Network Model

NARX neural network is a variant of Recurrent Network (Lin *et al.* 1996, Gao & Meng, 2005) that has been successfully

utilized in time series prediction problems. The major difference between RNN and MLFFN is that RNN allows a weighted feedback connection between layers of neurons and thereby making it suitable for time series analysis by allowing lagged values of variables to be considered in model. Although throughout the literature many time series methods such Autoregressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), etc. have been applied in various econometrics problems, these techniques cannot cope with nonlinear problems. NARX on the contrary can efficiently be used for modelling non stationary and nonlinear time series. Mathematically input output representation of nonlinear discrete time series in NARX network is governed by the following equation.

$$y(t) = f[u(t - D_u), \ldots\ldots, u(t-1), u(t), y(t - D_y), \ldots\ldots, y(t-1)] \quad (5)$$

where u(t) and g(t) represent input and output of the network at time t. $D_u$ and $D_y$, are the input and output order, and the function f is a nonlinear function. The function is approximated by MLFFN. It is also possible to have NARX networks with zero input order and a one-dimensional output. i.e., having feedback from output only. In such cases operation of NARX network is governed by equation 6.

$$y(t) = \Psi[u(t), y(t-1), \ldots\ldots y(t-D)] \quad (6)$$

where $\Psi$ is the mapping function, approximated by standard MLFFNN. Schematic structure of NARX is depicted in figure 2.

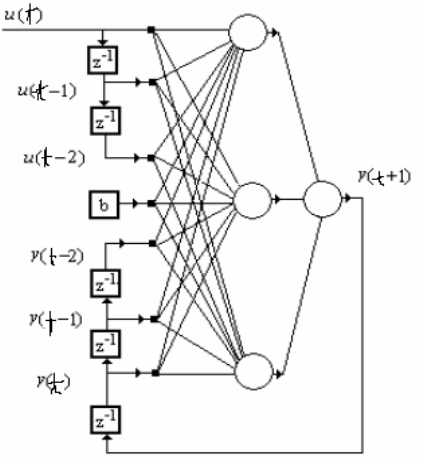


**Figure 2.** *NARX architecture*

## Time Series Modelling

Objective of time series forecasting is to predict future values of a time series, $X_{n+m}$ based on the observed data to present, $X = \{X_n, X_{n-1},\ldots..,X_1\}$. Majority of the time series forecasting model assume $X_t$ to be stationary. Autoregressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), Autoregressive Distributed Lag (ARDL), etc. are various time series modelling techniques that are predominantly applied to forecast linear and univariate time series. Nonstationary time series can be converted to stationary by various means to fit these models. However it has been observed that financial data exhibits volatility clustering i.e., high volatile periods are followed by high volatile periods and low volatile periods by low volatile periods. One of the major limitation of the above mentioned models is that they are all built under the assumption that conditional variance of past is constant. For heteroskedastic situations traditional linear time series forecasting techniques fail to capture the volatility and thereby yield poor predictions. Autoregressive Conditional

Heteroskedastic (ARCH) proposed by Engle (1982) was introduced to model volatility. It has been further extended to Generalized Autoregressive Conditional Heteroskedastic (GARCH) model by Bollerslev (1986), Exponential GARCH (EGARCH) by Nelson (1990), Threshold ARCH (TARCH) by Rabemananjara and Zakoian in 1993, Quadratic ARCH by Sentana in 1995, etc. In this study, we have applied multivariate GARCH and EGARCH model. Generalized description of both these two models are furnished below.

## GARCH

GARCH (p, q) model is actually same as ARCH model of infinite order. For GARCH (p, q) model, the conditional volatility ($h_t$) is a function of previous conditional volatility ($h_{t-p}$) and previous squared error ($\varepsilon_{t-q}^2$). The standard GARCH (1, 1) for stock returns model can be represented by following equations.

$$R = c + \rho R_{t-1} + \varepsilon_t \quad (7)$$

$$\varepsilon_t = z_t\sqrt{h_t} \quad (8)$$

Where $z_t \sim N(0,1)$ and

$$h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1} \quad (9)$$

All the parameters are positive and $(\alpha + \beta)$ measures the persistence of volatility. In general $(\alpha + \beta)<1$ and has been observed to be very close to 1. The effect of any shock in volatility decays at a rate of$(1 - \alpha - \beta)$.

In case of GARCH (p, q) model equation 9 becomes

$$h_t = \omega + \alpha_1\varepsilon_{t-1}^2 + \alpha_2\varepsilon_{t-2}^2 + \cdots + \alpha_q\varepsilon_{t-q}^2 + \beta_1 h_{t-1} + \beta_2 h_{t-2} + \cdots + \beta_p h_{t-p}$$

$$= \omega + \sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}^2 + \sum_{i=1}^{p}\beta_i h_{t-i} \quad (10)$$

Since our research framework is multivariate in nature, we have utilized multivariate extension of GARCH model in our research.

### EGARCH

It is a variant of GARCH model. Formal EGARCH (1, 1) model can be characterized by:

$$\log(h_t) = \omega + \beta \log(h_{t-1}) + \alpha \left| \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} \right| + \gamma \frac{\varepsilon_{t-1}}{h_{t-1}} \qquad (11)$$

The parameter $\alpha$ measures the magnitude of volatility clustering. As the conditional variance is measured in logarithmic form, it allows the coefficients to have negative values. The parameter $\gamma$ captures the leverage effect.

### Unit Root Test

As the time series must be stationary for GARCH, EGARCH modelling, Augmented Dickey Fuller (ADF) test is conducted to check the presence of unit roots. For a univariate time series, $y_t$ the ADF test basically applies regression to the following model:

$$\Delta y_t = \alpha + \beta_t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \cdots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t \qquad (12)$$

Where $\alpha$, $\beta$ and $p$ are constant, coefficient on a time trend and lag order of autoregressive process. It corresponds to random walk model if $\alpha = 0$ and $\beta = 0$ constraints are imposed. Whereas using the constraint $\beta = 0$ corresponds to modelling random walk with a drift. The Unit Root Test is then conducted under null hypothesis ($H_0$) $\gamma = 0$ against alternative hypothesis ($H_1$) $\gamma < 0$. The acceptance of null hypothesis implies nonstationary.

To quantitatively judge the performance of these models Mean Squared Error (MSE), Correlation Coefficient (R) and Theil Inequality (TI) measures are obtained. They are computed using the following set of equations.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\{Y_{act}(i)-Y_{pred}(i)\}^2 \quad (13)$$

$$R = \frac{\sum_{i=1}^{N}(Y_{act}(i)-\overline{Y_{act}})(Y_{pred}(i)-\overline{Y_{pred}})}{\sqrt{\sum_{i=1}^{N}(Y_{act}(i)-\overline{Y_{act}})^2}\sqrt{\sum_{i=1}^{N}(Y_{pred}(i)-\overline{Y_{pred}})^2}} \quad (14)$$

$$TI = \frac{\left[\frac{1}{N}\sum_{i=1}^{N}\left(Y_{act}(i)-Y_{pred}(i)\right)^2\right]^{1/2}}{\left[\frac{1}{N}\sum_{i=1}^{N}Y_{act}(i)^2\right]^{1/2}+\left[\frac{1}{N}\sum_{i=1}^{N}Y_{pred}(i)^2\right]^{1/2}} \quad (15)$$

Where $Y_{act}(i)$ and $Y_{pred}(i)$ are actual observed and predicted value of $i^{th}$ sample. N is the sample size. Whereas $\overline{Y_{act}}$ and $\overline{Y_{pred}}$ denote the average of actual and predicted values of N samples. Values MSE must be as low as possible for efficient prediction; ideally a value of zero signifies no error or perfect prediction. Both R and TI Values range between [0, 1]. They should be close to 1 for good prediction while 0 implies no prediction at all.

## Results of ann based modeling

As discussed we have utilized two different ANN models, MLFFNN, a traditional model, and NARX, a tailor made tool for time series modelling. The entire dataset from 1.1.2009 to 8.4.2016 has been suitably partitioned into training, validation and test dataset (70%, 15% & 15%) for predictive modelling exercise. Performance of the respective models are evaluated using Mean Squared Error (MSE) and Correlation Coefficient (R) measures

### Results of MLFFNN modelling

Only one hidden layer has been used while number of neurons in hidden layer has been varied at four levels (10,

20, 30 & 40 number of neurons). Levenberg-Marquardt (LM), Scaled Conjugate Gradient (SCG), Conjugate Gradient with Powell-Beale Restarts (CGPB), Fletcher-Powell Conjugate Gradient (FPCG), Polak-Ribiére Conjugate Gradient (PRCG) as five different variants of backpropagation algorithms have been adopted for learning purpose. So total twenty (no. of levels of neurons × no. of learning algorithms) numbers of experimental trials are conducted. Other specifications are mentioned in the following table.

**Table 3.** *MLFFNN parameter settings*

| Sl. No. | Parameter | Data/Technique Used |
|---|---|---|
| 1. | Number of input neuron(s) | Eight |
| 2. | Number of output neuron(s) | One |
| 3. | Transfer function(s) | Tan-sigmoid transfer function (tansig) in hidden layer &purelin in output layer. |
| 4. | Proportion of training, validation and test dataset | 70:15:15 |
| 5. | Error function(s) | Mean squared error(MSE) function |
| 6. | Type of Learning rule | Supervised learning rule |

The following figure depicts the strength of association between the target and output (predicted) exchange rate on training, validation, testing and entire dataset. Correlation coefficient values on respective data have also been mentioned.

One sample MLFFNN structure out of 20 trials is shown below.

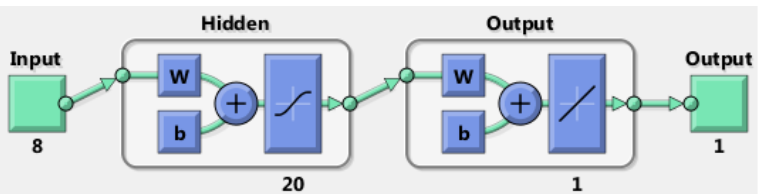


**Figure 3.** *Sample MLFFNN Structure*

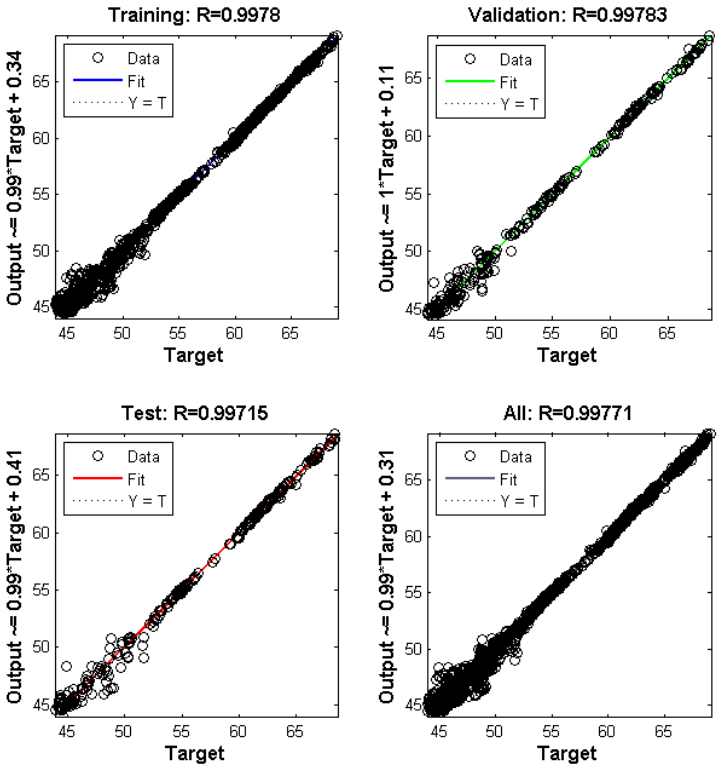


**Figure 4.** *Performance of MLFFNN*

It can be clearly seen that actual exchange rate (represented by target in the graph) and predicted exchange rate (represented by output in the graph) are very close for training, validation, test and entire data set. Almost a linear trend can be observed between the actual and predicted exchange rate which justifies the efficacy of the model. To further validate the claim MSE value is also obtained for individual experimental trials.

Statistics of MSE and R of predictive modelling on training and test dataset are summarized in following tables. For computation of MSE actual and predicted values have been rescaled to [0, 1] for all models.

**Table 4.** *Performance on Training Dataset*

| Statistics | R | MSE |
|---|---|---|
| Min | 0.9916 | 0.000132 |
| Max | 0.9989 | 0.000364 |
| Average | 0.9974 | 0.000227 |

**Table 5.** *Performance on Test Dataset*

| Statistics | R | MSE |
|---|---|---|
| Min | 0.9907 | 0.000169 |
| Max | 0.9981 | 0.000417 |
| Average | 0.9957 | 0.000289 |

High R values and negligible MSE values for both training and test data set imply that exchange rate can effectively be predicted using MLFFNN architecture using FX4, DJIAR, NIFTYR, DR, HSR, COP, CV and IV.

## Results of NARX modelling

Similar to MLFFNN, only one hidden layer is employed in NARX network too. Delay of 2 units to consider the lagged values of both dependent and independent variables have been considered for model building. Number of neurons in hidden layer is varied at four levels and five learning algorithms have been used.

For the considered problem, general formulation of NARX structure as indicated in equation 6 is replaced by equation 16.

$$\text{FX(t)} = f(FX4(t), FX4(t-1), FX4(t-2), NIFTY(t), NIFTY(t-1), NIFTY(t-2), DJIAR(t), DJIAR(t-1), DJIAR(t-2), HSR(t), HSR(t-1), HSR(t-2), DR(t), DR(t-1), DR(t-2), COP(t), COP(t-1), COP(t-2), CV(t), CV(t-1), CV(t-2), IV(t), IV(t-1), IV(t-2)) \quad (16)$$

Similar to MLFFNN, five backpropagation algorithms namely, Levenberg-Marquardt (LM), Scaled Conjugate

Gradient (SCG), Conjugate Gradient with Powell-Beale Restarts (CGB), Fletcher-Powell Conjugate Gradient (CGF), Polak-Ribiére Conjugate Gradient (CGP) are used for training. Other specifications of NARX are same as of MLFNN highlighted in Table 3. One sample NARX structure is displayed in Figure 5.

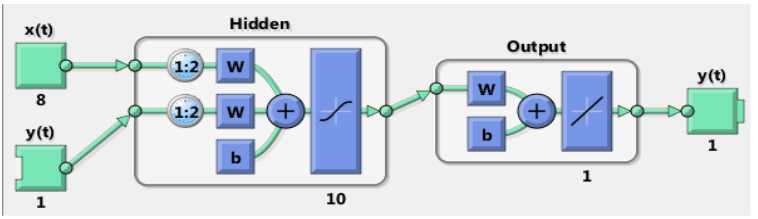


**Figure 5.** *Sample NARX structure*

Figure 6 graphically represents association of the actual and predicted exchange rate for all training, test and validation sample. Magnitude of error expressed as difference between actual and predicted values is also shown in the same figure. Although visual representation strongly suggests goodness of fit of NARX network in predicting exchange rate, to quantitatively justify the claim, MSE and R values are computed for training and test dataset.

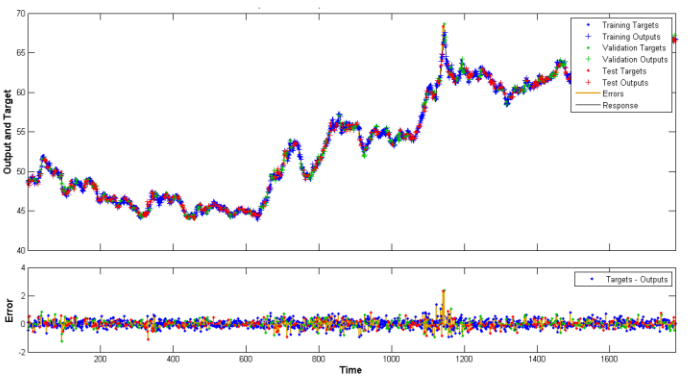


**Figure 6.** *Visualization of NARX performance*

Statistics of MSE and R of predictive modelling performance of NARX network on training and test dataset for all experimental trials are summarized in following tables.

**Table 6.** *Performance on Training Dataset*

| Statistics | R | MSE |
|---|---|---|
| Min | 0.9911 | 0.000125 |
| Max | 0.9979 | 0.000373 |
| Average | 0.9957 | 0.000229 |

**Table 7.** *Performance on Test Dataset*

| Statistics | R | MSE |
|---|---|---|
| Min | 0.9898 | 0.000169 |
| Max | 0.9942 | 0.000392 |
| Average | 0.9923 | 0.000271 |

It is evident from negligible MSE and high R values that the presented NARX network with 2 delay units has predicted exchange rate as a nonlinear function of FX4, DJIAR, NIFTYR, HSR, DR, COP, CV and IVquite effectively. Error histogram with 20 bins is displayed below.

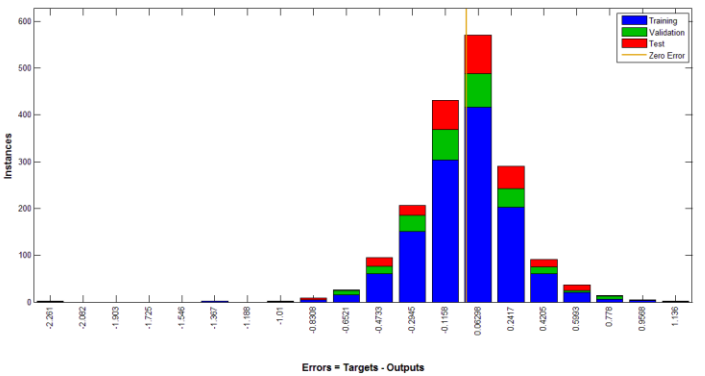


**Figure 7.** *Error histogram of NARX modelling*

## Assesment of parametres

To determine the impact of number of neurons in hidden layer and different backpropagation algorithms on performance in terms of MSE of MLFFNN and NARX on test data set, Analysis of Covariance (ANCOVA) has been performed. Different algorithms and number of neurons are treated as fixed factor and covariate respectively. Results are summarized in Table 8.

**Table 8.** *Tests of Between-Subjects Effects (MLFFNN)*

| Dependent Variable: MSE | | | | | | |
|---|---|---|---|---|---|---|
| Source | Type III Sum of Squares | Df | Mean Square | F | Sig. | Partial Eta Squared |
| Corrected Model | 3.662E-008 | 5 | 7.324E-009 | 2.010 | .139 | .418 |
| Intercept | 1.644E-007 | 1 | 1.644E-007 | 45.130 | .000 | .763 |
| Neurons | 9.604E-011 | 1 | 9.604E-011 | .026 | .873 | .002 |
| Algorithms | 3.652E-008 | 4 | 9.131E-009 | 2.506 | .089 | .417 |
| Error | 5.101E-008 | 14 | 3.643E-009 | | | |
| Total | 1.118E-006 | 20 | | | | |
| Corrected Total | 8.763E-008 | 19 | | | | |

It is observed that varying the number of neurons in hidden layer does not have significant impact on predictive performance of MLFFNN. On the other hand, usage of different backpropagation algorithms has somewhat influence (at p-value < 0.1 level) on the performance. It can be inferred that number of neurons in hidden layer can be fixed at any of the four levels considered in this. The same proposition cannot be made for the various back propagation algorithms deployed though. In future, further investigations can be made using advanced Taguchi's experimental design methods or Response Surface Methodology to find the optimum level parameter settings.

**Table 10.** *Tests of Between-Subjects Effects (NARX)*

| Dependent Variable: MSE | | | | | | |
|---|---|---|---|---|---|---|
| Source | Type III Sum of Squares | Df | Mean Square | F | Sig. | Partial Eta Squared |
| Corrected Model | 5.731E-009 | 5 | 1.146E-009 | .202 | .956 | .067 |
| Intercept | 3.032E-007 | 1 | 3.032E-007 | 53.549 | .000 | .793 |
| Neurons | 3.745E-009 | 1 | 3.745E-009 | .661 | .430 | .045 |
| Algorithms | 1.986E-009 | 4 | 4.964E-010 | .088 | .985 | .024 |
| Error | 7.927E-008 | 14 | 5.662E-009 | | | |
| Total | 1.554E-006 | 20 | | | | |
| Corrected Total | 8.500E-008 | 19 | | | | |

For NARX model neither varying the number of neurons nor the usage of different training algorithms significantly affect the overall performance. Hence any specifications of parameters of NARX model out of twenty experimental setups can be suitably chosen for prediction of exchange rate.

## Results of time series based modeling

As discussed, before proceeding with GARCH and EGARCH modelling to check whether the dataset is stationary or not, ADF test, discussed in section 3, is conducted. Additionally Philips-Perron (PP) test has been conducted too for the same. Similar to ADF test, acceptance of null hypothesis in PP test means the time series is nonstationary. Results are summarized in following table.

**Table 11.** *Results of Unit Root Test (ADF)*

| Variable | t-Statistic | p-value | Significant |
|---|---|---|---|
| FX1 | -0.16374 | 0.9404 | Not Significant |
| FX4 | -0.33843 | 0.9167 | Not Significant |
| NIFTYR | -39.82233 | 0.0000 | *** |
| DJIAR | -45.30706 | 0.0001 | *** |
| HSR | -42.06761 | 0.0000 | *** |
| DR | -41.16089 | 0.0000 | *** |
| COP | -42.82212 | 0.0000 | *** |
| CV | -4.221728 | 0.0006 | *** |
| IV | -3.676932 | 0.0005 | *** |

*** Significant at 1% level.

**Table 12.** *Results of Unit Root Test (Philips Perron Test)*

| Variable | Adj. t-Statistic | p-value | Significant |
|---|---|---|---|
| FX1 | -0.33662 | 0.917 | Not Significant |
| FX4 | -0.43354 | 0.901 | Not Significant |
| NIFTYR | -39.7629 | 0.0000 | *** |
| DJIAR | -45.304 | 0.0001 | *** |
| HSR | -42.0675 | 0.0000 | *** |
| DR | -41.1526 | 0.0000 | *** |
| COP | -42.8303 | 0.0000 | *** |
| CV | -13.9375 | 0.0000 | *** |
| IV | -4.53662 | 0.0002 | *** |

*** Significant at 1% level

Both ADF and PP tests suggest that FX1 and FX4 are nonstationary. As only FX1 and FX4 are found to be nonstationary, we have taken the first order difference of these two variables and further applied ADF test to check the stationary constraint before building GARCH and EGARCH models.

**Table 13.** *Results of Unit Root Test (First Difference Series) via ADF*

| Variable | t-statistic | p-value | Significance |
|---|---|---|---|
| FX1 | -31.87125 | 0.0000 | *** |
| FX4 | -32.81953 | 0.0000 | *** |

*** Significant at 1% level.

**Table 14.** *Results of Unit Root Test (First Difference Series) via Philips Perron Test*

| Variable | Adj. t-Statistic | p-value | Significance |
|---|---|---|---|
| FX1 | -40.5455 | 0.0000 | *** |
| FX4 | -41.0302 | 0.0000 | *** |

*** Significant at 1% level.

From above two tables, FX1 and FX4 are identified as I(1). Subsequently ARCH Lagrange Multiplier (LM) test is performed. LM test statistic values and corresponding p-values for different duration of lags are reported in Table 15.

**Table 15.** *Results of ARCH LM Test*

| Lag | F-statistic | p-value | Significance |
|---|---|---|---|
| 1-2 | 3077.471 | 0.0000 | *** |
| 1-5 | 1270.281 | 0.0000 | *** |
| 1-10 | 639.3180 | 0.0000 | *** |
| 1-15 | 425.0291 | 0.0000 | *** |

*** Significant at 1% level

As LM test statistic values are significant, presence the ARCH effect is deduced. These findings justifies the incorporation of GARCH and EGARCH in this research problem. We have utilized GARCH (1, 1), GARCH (2, 2), EGARCH (1, 1) and EGARCH (2, 2) models for forecasting purpose. Model fitness in terms of R-squared, Adjusted R-squared, Akaike Information Criterion (AIC), Schwarz Information Criterion (SIC) and Hannan-Quinn Information Criterion (HQC) are calculated for respective models and mentioned in Table 16 and 17.

**Table 16.** *Results of GARCH Model*

| Model | R-squared | Adjusted R-squared | AIC | SC | HQC |
|---|---|---|---|---|---|
| GARCH(1,1) | 0.971979 | 0.971853 | -0.299478 | -0.262522 | -0.285829 |
| GARCH(2,2) | 0.973544 | 0.973424 | -0.292701 | -0.249585 | -0.276777 |

**Table 17.** *Results of EGARCH Model*

| Model | R-squared | Adjusted R-squared | AIC | SC | HQC |
|---|---|---|---|---|---|
| EGARCH(1,1) | 0.973361 | 0.973241 | -0.249598 | -0.209563 | -0.234812 |
| EGARCH(2,2) | 0.971791 | 0.971663 | -0.309553 | -0.263357 | -0.292491 |

As the values of the critical model indices are quite good, usage of volatility models for forecasting exchange rate is well justified. Subsequently estimated coefficient values of predictor variables by all four employed models are serially reported.

**Table 18.** *Estimated Parameters of GARCH (1, 1) Model*

| Variable | Coefficient | Std. Error | z-Statistic | p-Value | Significance |
|---|---|---|---|---|---|
| Intercept/Constant | 0.133896 | 0.015425 | 8.680527 | 0.00000 | *** |
| FX4 | 0.979148 | 0.000192 | 5106.78 | 0.00000 | *** |
| NIFTYR | -0.26702 | 0.153306 | -1.74172 | 0.00000 | *** |
| DJIAR | 0.587357 | 0.186016 | 3.157555 | 0.00000 | *** |
| HSR | 0.004415 | 0.109882 | 0.040177 | 0.968 | Not Significant |
| DR | 0.170601 | 0.115334 | 1.479195 | 0.1391 | Not Significant |
| COP | -0.03279 | 0.09851 | -0.33282 | 0.7393 | Not Significant |
| CV | 0.003903 | 0.000257 | 15.15883 | 0.00000 | *** |
| IV | 0.00138 | 0.000315 | 4.377546 | 0.00000 | *** |

*** Significant at 1% level.

Results reveal that out of eight predictors used in mean equation of GARCH model, FX4, NIFTYR, DJIAR, CV and IV are statistically significant.

**Table 19.** *Estimated Parameters of GARCH (2, 2) Model*

| Variable | Coefficient | Std. Error | z-Statistic | p-Value | Significance |
|---|---|---|---|---|---|
| Intercept/Constant | -0.13044 | 0.012862 | -10.1412 | 0.00000 | *** |
| FX4 | 0.981626 | 0.000174 | 5650.204 | 0.00000 | *** |
| NIFTYR | 0.499303 | 0.138783 | 3.59772 | 0.0002 | *** |
| DJIAR | 0.220602 | 0.205063 | 1.075773 | 0.282 | Not Significant |
| HSR | -0.00017 | 0.108063 | -0.00156 | 0.9988 | Not Significant |
| DR | 0.149937 | 0.12472 | 1.202196 | 0.2293 | Not Significant |
| COP | -0.14303 | 0.083851 | -1.70572 | 0.0881 | * |
| CV | 0.004391 | 0.000244 | 18.0167 | 0.00000 | *** |
| IV | 0.00779 | 0.000246 | 31.68688 | 0.00000 | *** |

*** Significant at 1% level. * Significant at 10% level.

For GARCH (2, 2) model FX4, NIFTYR, CV and IV are found to be highly significant. COP is significant at 10% level. Unlike GARCH (1, 1), DJIAR has been marked as not significant. HSR and DR are not significant as well.

**Table 20.** *Estimated Parameters of EGARCH (1, 1) Model*

| Variable | Coefficient | Std. Error | z-Statistic | p-Value | Significance |
|---|---|---|---|---|---|
| Intercept/Constant | -0.12546 | 0.009768 | -12.8437 | 0.00000 | *** |
| FX4 | 0.981681 | 0.000118 | 8349.087 | 0.00000 | *** |
| NIFTYR | 0.43457 | 0.117823 | 3.688328 | 0.00000 | *** |
| DJIAR | 0.178091 | 0.168117 | 1.05933 | 0.2894 | Not Significant |
| HSR | 0.01102 | 0.113189 | 0.097361 | 0.9224 | Not Significant |
| DR | 0.103022 | 0.113425 | 0.908286 | 0.3637 | Not Significant |
| COP | 0.009301 | 0.071747 | 0.129633 | 0.8969 | Not Significant |
| CV | 0.003717 | 0.000219 | 16.97872 | 0.00000 | *** |
| IV | 0.007858 | 0.000267 | 29.39315 | 0.00000 | *** |

*** Significant at 1% level.

In EGARCH (1, 1) model, significant predictors are turned out to be FX4, NIFTYR, CV and IV. Rest four predictors does not have significant impact on exchange rate.

**Table 21.** *Estimated Parameters of EGARCH (2, 2) Model*

| Variable | Coefficient | Std. Error | z-Statistic | p-Value | Significance |
|---|---|---|---|---|---|
| Intercept/Constant | 0.167494 | 0.014549 | 11.51244 | 0.00000 | *** |
| FX4 | 0.978905 | 0.000173 | 5659.589 | 0.00000 | *** |
| NIFTYR | -0.25991 | 0.142216 | -1.82758 | 0.0676 | * |
| DJIAR | 0.60209 | 0.153425 | 3.924324 | 0.0001 | *** |
| HSR | 0.085213 | 0.091592 | 0.930357 | 0.3522 | Not Significant |
| DR | 0.282888 | 0.098929 | 2.859516 | 0.0032 | *** |
| COP | -0.02665 | 0.066709 | -0.39949 | 0.6895 | Not Significant |
| CV | 0.003608 | 0.000284 | 12.69635 | 0.00000 | *** |
| IV | 0.000697 | 0.000284 | 2.449421 | 0.0143 | ** |

*** Significant at 1% level, ** Significant at 5% level, * Significant at 1% level.

Apart from HSR and COP, rest six independent variables have significant impact on movement of exchange rate according to EGARCH (2, 2) model.

To visualize the forecasting results obtained from GARCH (1, 1), GARCH (2, 2), EGARCH (1, 1) and EGARCH (2, 2) the following figures are presented.

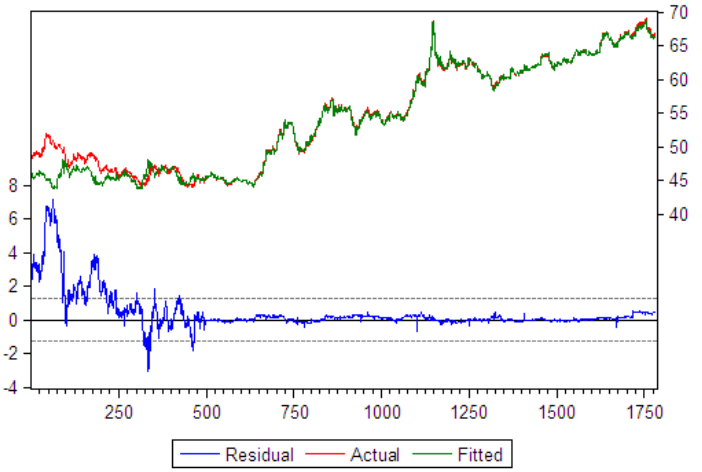


**Figure 8.** *GARCH (1, 1) Performance*

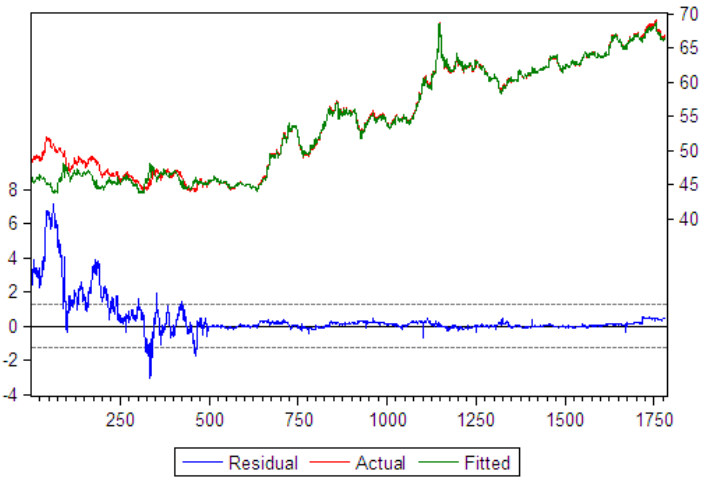


**Figure 9.** *GARCH (2, 2) Performance*

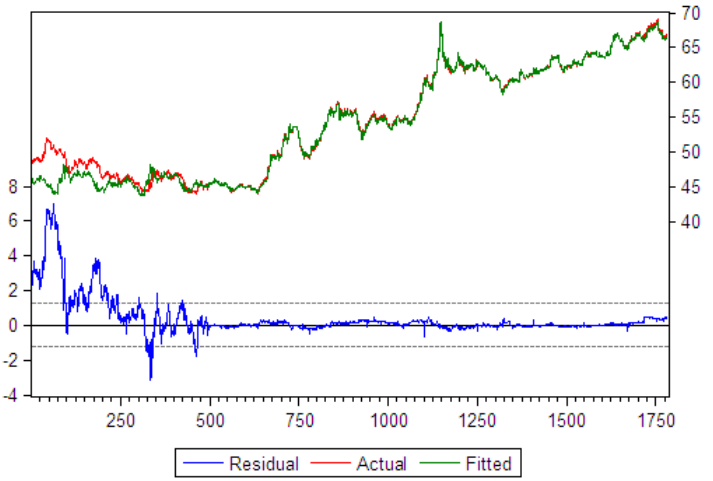


**Figure 10.** *EGARCH (1, 1) Performance*

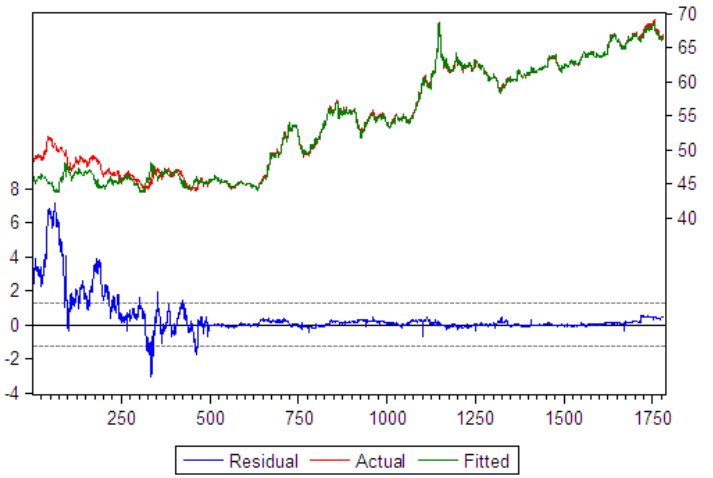


**Figure 11.** *EGARCH (2, 2) Performance*

Statistics of residuals are presented in Table 22. Conditional heteroscedasticityof residuals can be observed

in terms of Kurtosis and Jarque-Bera Statistic that strongly justifies effectiveness of utilized time series framework.

**Table 22.** *Residual Diagnostic*

| Model | Mean | Median | Maximum | Minimum | Std. Dev. | Skewness | Kurtosis | Jarque-Bera | p-value |
|---|---|---|---|---|---|---|---|---|---|
| GARCH (1, 1) | 0.2809 | 0.7071 | 4.2983 | -5.4856 | 0.9600 | -0.6924 | 4.1962 | 248.4826 | 0.0000 |
| GARCH (2, 2) | 0.2779 | 0.7022 | 4.2727 | -5.1774 | 0.9610 | -0.7071 | 4.0186 | 225.4190 | 0.0000 |
| EGARCH (1, 1) | 0.2544 | 0.6309 | 4.5352 | -5.0639 | 0.9674 | -0.6969 | 4.1984 | 250.7210 | 0.0000 |
| EGARCH (2, 2) | 0.2788 | 0.6888 | 3.7100 | -5.4237 | 0.9597 | -0.6967 | 3.4505 | 159.1212 | 0.0000 |

Lastly for quantitative assessment of forecasting accuracy, MSE and Theil Inequality Coefficient are calculated using actual and obtained forecast values for all samples and shown in Table 23.

**Table 23.** *Forecasting Performance*

| Model | MSE | Theil Inequality Coefficient |
|---|---|---|
| GARCH (1, 1) | 0.002310165 | 0.01152 |
| GARCH (2, 2) | 0.007260015 | 0.01550 |
| EGARCH (1, 1) | 0.00223638 | 0.01123 |
| EGARCH (2, 2) | 0.00235386 | 0.01552 |

Since both MSE and Theil Inequality Coefficient are substantially low, conclusions can be drawn that all the four models have been quite effective for forecasting exchange rate. Performance both GARCH and EGARCH model in multivariate framework is well justified.

## Comparative analysis

In terms of MSE measures, performance of four GARCH family models and two ANN models are graphically plotted in figure 12.

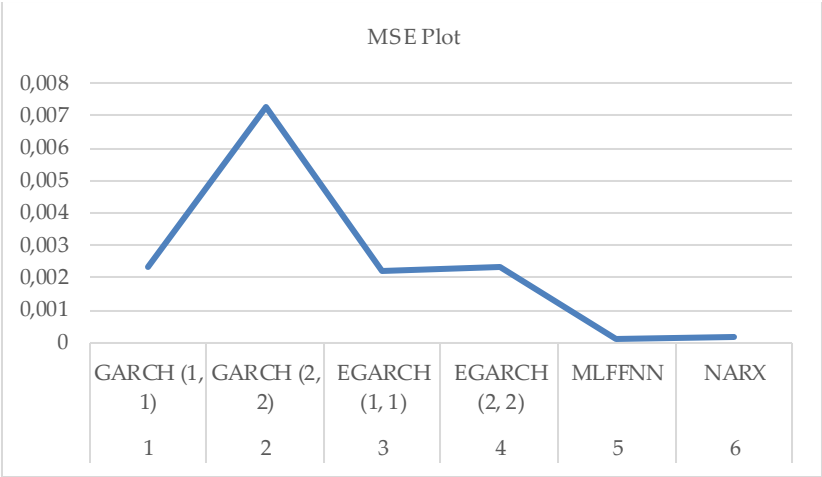


**Figure 12.** *Comparative Analysis*

Graphically it is quite evident that both the ANN models perform better than the four GARCH models. To statistically justify the claim t-test has been performed on MSE to determine whether the performance of ANN and GARCH family models are significantly different or not.

**Table 24.** *t-Test Result (on test cases)*

| p-Value | Significance |
|---|---|
| 0.0127 | ** |

** Significant at 5% level

As the t-Test statistic is significant at 5% level, it can be concluded that there is a significant difference between the performance of ANN and GARCH family models in predicting the exchange rates. ANN models are better than GARCH models in terms of MSE values.

## Concluding remark

The chapter uses both ANN based models and Econometric models in a multivariate framework to predict the Indian rupee US dollar exchange rate. The study is based

on daily data. It incorporates explanatory variables from both the current account and the capital account of the balance of payments. During the process of generating results, it was observed that both sets of techniques generated useful and efficient predictions of the exchange rate. Further, the explanatory variables chosen were quite appropriate for the study. The application of both MLFFNN and NARX including the use of various backpropagation algorithms is quite unique and the non-linear relationship between the exchange rate and the explanatory variables have been effectively captured.

From the technique point of view, it is observed that the predictive performance of MLFFNN does not depend on the number of neurons in the hidden layer, but is sensitive to the backpropagation algorithms. For the NARX model, neither the number of neurons, nor the training algorithms, significantly affect the performance.

In econometric modelling, four different approaches namely, GARCH (1,1,), GARCH (2,2), EGARCH (1,1) and EGARCH (2,2) were used and the results have been reported. While the results obtained have been satisfactory, a comparative analysis of the ANN based models and the econometric models reveals that MLFFNN and NARX are better methods in terms of predictive efficiency.

## References

Andreou, A.S., & Zombanakis, G.A. (2006). Computational intelligence in exchange-rate forecasting, *Bank of Greece Working Paper*, No.49.

Bhat, S.A., & Nain, Md.Z. (2014). Modelling the conditional heteroscedasticity and leverage effect in the BSE sectoral indices, *IUP Journal of Financial Risk Management*, 11, 49-61.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 31(3), 307-327. doi. 10.1016/0304-4076(86)90063-1

Caves, R.E. & Jones, R.W. (1981). *World Trade and Payments – An Introduction*, 3rd Edition, Little Brown and Company, Boston, Toronto.

Datta Chaudhuri, T., & Ghosh, I. (2015). Forecasting volatility in Indian stock market using artificial neural network with multiple inputs and outputs, *International Journal of Computer Applications*, 120(8), 7-15. doi. 10.5120/21245-4034

Dua, P., & Ranjan, R. (2011). Modelling and forecasting the Indian Re/Us Dollar exchange rate, *RBI Working Paper*, No.197.

Engle, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, 50(4), 987-1008. doi. 10.2307/1912773

Gao, Y., & Meng, J.E. (2005). NARMAX time series model prediction: feedforward and recurrent fuzzy neural network approaches, *Fuzzy Sets and Systems*, 150(2), 331-350. doi. 10.1016/j.fss.2004.09.015

Garg, A. (2012). Forecasting exchange rates using machine learning models with time-varying volatility, *Master Thesis*, [Retrieved from].

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques*, 3rd Edition, Morgan Kauffman Publishers, USA.

Haykin, S. (1999). *Neural Networks*, Second Edition, Pearson Education.

Imam, T. (2012). Intelligent computing and foreign exchange rate prediction: What we know and we don't, *Progress in Intelligent Computing and Applications*, 1(1), 1-15. doi. 10.4156/pica.vol1.issue1.1

Jena, P.R., Majhi, R., & Majhi, B. (2015). Development and performance evaluation of a novel knowledge guided artificial neural network (KGANN) model for exchange rate prediction, *Journal of King Saud University - Computer and Information Sciences*, 27(4), 450-457. doi. 10.1016/j.jksuci.2015.01.002

Lam, L., Fung, L., & Yu, I.W. (2008). Comparing forecast performance of exchange rate models, *Hong Kong Monetary Authority Working Paper*, No.08.

Lin, T., Horne, B.G., Tino, P., & Giles, C.L. (1996). Learning long-term dependencies in NARX recurrent neural networks, *IEEE Transactions on Neural Networks*, 7(6), 1329-1351. doi. 10.1109/72.548162

Lin, S.Y., Chen, C.H., & Lo, C.C. (2013). Currency exchange rates prediction based on linear regression analysis using cloud computing, *International Journal of Grid and Distributed Computing*, 6, 1-10.

Majhi, B., Rout, M., Majhi, R., Panda, G., & Fleming, P.J. (2012). New robust forecasting models for exchange rates prediction, *Expert Systems with Applications*, 39(16), 12658 - 12670. doi. 10.1016/j.eswa.2012.05.017

Messe, R.A., & Rogoff, K.K. (1983). Empirical exchange rate models of the seventies do they fit out of samples, *Journal of International Economics*, 14(1-2), 3-24. doi. 10.1016/0022-1996(83)90017-X

Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: A new approach, *Econometrica*, 59(2), 347–370. doi. 10.2307/2938260

Pacelli, V. (2012). Forecasting exchange rates: A comparative analysis, *International Journal of Business and Social Science*, 3(10), 145-156.

Perwej, Y., & Perwej, A. (2012). Forecasting of Indian Rupee (INR) / US Dollar (USD) currency exchange rate using artificial neural network, *International Journal of Computer Science, Engineering and Applications*, 2(2), 41-52. doi. 10.5121/ijcsea.2012.2204

Premanode, B., & Toumazou, C. (2013). Improving prediction of exchange rates using Differential EMD, *Expert Systems with Applications*, 40(1), 377-384. doi. 10.1016/j.eswa.2012.07.048

Rabemananjara, R., & Zakoian, J.M. (1993). Threshold ARCH models and asymmetries in volatility, *Journal of Applied Econometrics*, 8(1), 31–49. doi. 10.1002/jae.3950080104

Ravi, V., Lal, R., & Kiran, N.R. (2012). Foreign exchange rate prediction using computational intelligence Methods, *International Journal of Computer Information Systems and Industrial Management Applications*, 4, 659-670.

Sentana, E. (1995). Quadratic ARCH models, *Review of Economic Studies*, 62(4), 639–661. doi. 10.2307/2298081

Tripathy, S., & Rahaman, A. (2013). Forecasting daily stock volatility using GARCH model: A comparison between BSE and SSE, *IUP Journal of Applied Finance*, 19, 71-83.

Vojinovic, Z., & Kecman, V. (2001). A data mining approach to financial time series modelling and forecasting, *Intelligent Systems in Accounting, Finance and Management*, 10(4), 225-239. doi. 10.1002/isaf.207

Zhang, G.P., & Berardi, V.L. (2001). Time series forecasting with neural network ensembles: An application for exchange rate prediction, 52, 652-664. doi. 10.1057/palgrave.jors.2601133

**For Cited this Chapter:**

**Chaudhuri, T.D., & Ghosh, I. (2020). Artificial neural network and time series modeling based approach to forecasting the exchange rate in a multivariate framework, *In* T.D. Chaudhuri (Ed.), *Overview of the Indian Economy: Empirical analysis of financial markets, exchange rate movements and industrial development* (pp.141-178), KSP Books: Istanbul.**

ISBN: 978-605-7736-93-2 (e-Book)
KSP Books 2020
© KSP Books 2020

**Copyrights**

Copyright for this Book is retained by the author(s), with first publication rights granted to the Book. This is an open-access Book distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by-nc/4.0 ).

# 7 A predictive analysis of the Indian FMCG sector using time series decomposition - based approach

*By*
Jaydip SEN
& Tamal DATTA CHAUDHURI

## Introduction

One of the most exciting challenges to the researchers working in the field of machine learning and analytics is developing an accurate and efficient forecasting framework for predicting stock prices. Researchers working in this field have proposed various technical, fundamental and statistical indicators for predicting stock prices. (Sen & Datta Chaudhuri, 2016a; Sen & Datta Chaudhuri, 2016b; Sen & Datta Chaudhuri, 2016c) proposed a novel approach towards portfolio diversification and prediction of stock prices. The authors argued that different sectors in an economy do not exhibit identical pattern of variations in their stock prices. Different sectors exhibit different trend patterns, different seasonal characteristics and also differ in the randomness in their time series. While on one side the *efficient market hypothesis* has focused on the randomness aspect of stock price movements,

on the other side, there are propositions to disprove the hypothesis delving into various fundamental characteristics of different stocks. We contend that besides the differences in the fundamental characteristics among stocks of different companies, performances of different stocks also have a lot to do with the sectors to which the stocks belong. Since each sector has its own set of factors influencing its behavior, the price movements of stocks belonging to different sectors are guided by these factors. The factors responsible for the phenomenal growth of the *information technology* (IT) sector in India are different from those which have made the metals sector in the country sluggish, or the FMCG sector grow at a slow pace. From the point of view of investors in the stock market, it is critical to identify these factors and analyze them effectively for optimal portfolio choice and also for churning of the portfolio.

In this chapter, we focus on the time series pattern of the FMCG sector in India in order to understand its distinguishing characteristics. We use the monthly time series index values of the Indian FMCG sector during the period January 2010 till December 2016 as per the Bombay Stock Exchange (BSE). We decompose the time series using R programming language. We, then, illustrate how the time series decomposition approach provides us with useful insights into various characteristics and properties of the FMCG sector time series. It is further demonstrated that a careful and deeper study of the trend, seasonal and random components values of the time series enables one to understand the growth pattern, the seasonal characteristics and the degree of randomness exhibited by the time series index values. We also propose an extensive framework for time series forecasting in which we present six different approaches of prediction of time series index values. We critically analyze the six approaches and also explain the reason why some methods perform better and produce

lower values of forecast error in comparison to other methods.

The rest of the chapter is organized as follows. Section 2 presents the detailed methodology used in this work. It provides the details of the method of decomposition of time series into its various components. Section 3 depicts the results of decomposition of the FMCG sector time series index values into its trend, seasonal and random components. Based on the decomposition results, several characteristics and behavior exhibited by the FMCG sector time series are also analyzed. Section 4 provides comprehensive details of six forecasting methods that are proposed in this work. Section 5 presents extensive results on the performance of the six forecasting methods on the FMCG sector time series data. A comparative analysis of the techniques is also provided on the basis of six different metrics of the forecasting techniques: maximum error, minimum error, mean error, standard deviation of error, the *root mean square error* (RMSE), and the ratio of the RMSE value and the mean index value. Section 6 presents a brief discussion on some of the existing work in the literature on time series forecasting with particular focus on the FMCG sector. Finally, Section 7 concludes the chapter.

## Methodology

The methodology followed in this work is discussed in this section. The programming language R has been used in all phases of the work: data management, data analysis and presentation of data analysis results. (Ihaka & Gentleman, 1996) gave a detailed description of various capabilities of R programming language in data management and data analysis work. R is an open source language with a very rich set of libraries having in-built functions that makes it one of the most effective tools in handling data analysis projects. For the current work, we have used the monthly index data

from the Bombay Stock Exchange (BSE) of India for the FMCG sector for the period January 2010 till December 2016. The monthly index values of the FMCG sector for the 7 years are then stored in a plain text (.txt) file. This plain text file now contains 84 index values corresponding to the 84 months in the 7 year period under our study. The text file is then read into an R data object using the *scan( )* function. The R data object is then converted into a time series object by applying the *ts( )* function with a frequency value of 12. The frequency value is chosen to be 12 so that the seasonality characteristics of the time series for each month can be analysed. The time series data object in R is then decomposed into its three components – trend, seasonal and random – using the *decompose( )* function which is defined in the TTR library in R environment. We plot the graphs of the FMCG time series data as well as its three components so that further analysis can be made on the behavior of the time series and its three components.

After carrying out a comprehensive analysis of the decomposition results of the time series of the FMCG sector, we propose six different approaches of forecasting of time series index values. In order to compute the forecast accuracy of each method, we build the forecast models using the FMCG time series data for the period January 2010 till December 2015, and apply the models to forecast time series index values for each month of the year 2016. Since the actual values of the time series for all months of 2016 are already available with us, we compute the error in forecasting using each method of forecast that we have proposed. Comparative analysis of the methods of forecasting is done based on several useful metrics and reasons for which a particular method performs better than the other methods for the FMCG sector time series are clearly analyzed. A detailed comparative analysis,

highlighting which method performs best under what situations and for what type of time series, is also presented.

Demonstrated how effectively time series decomposition approach can be utilized in robust analysis and forecasting of the Indian Auto sector. In another different work, (Sen & Datta Chaudhuri, 2016a; Sen & Datta Chaudhuri 2016b) analyzed the behavior of two different sectors of Indian economy – the small cap sector and the capital goods sector – the former having a dominant random component while the latter exhibiting a significant seasonal component. Following another approach of time series analysis, (Sen & Datta Chaudhuri, 2016d) studied the behavior of the Indian *information technology* (IT) sector time series and the Indian *capital goods* sector time series. In yet another work, using the time series decomposition-based approach, (Sen & Datta Chaudhuri, 2016e) illustrated how time series analysis enables us to check the consistency between the fund style and actual fund composition of a mutual fund.

In this work, we demonstrate how time series decomposition-based approach enables one in analysing and understanding the behavior and different properties of the FMCG time series of the Indian economy based on time series data for the period January 2010 till December 2016. We also investigate which forecasting approach is most effective for the FMCG time series. For this purpose, we compare several approaches of forecasting and identify the one that produces the minimum value of forecasting error. We critically analyze all the proposed forecasting approaches, and explain why a particular approach has worked most effectively while some others have not done so for the FMCG time series data.

## Time series decomposition results

We present the decomposition results for the time series of the FMCG sector index values as per the records of the

BSE for the period January 2010 till December 2016. First, we create a plain text (.txt) file containing the monthly index values of the FMCG sector for the period January 2010 till December 2016. This file contained 84 records corresponding to the 84 months in the 7 years under study. We used the *scan( )* function in R language to read the text file and stored it in an R data object. Then, we converted this R data object into a time series object using the R function *ts( )*. We used the value of the *frequency* parameter in the *ts ( )* function as 12 so that the decomposition of the time series is carried out on monthly basis. After creating the time series data object, we used the function *plot( )* in R to draw the plot of the FMCG sector time series during the period January 2010 till December 2016. Figure 1 depicts the graph of the FMCG sector time series.

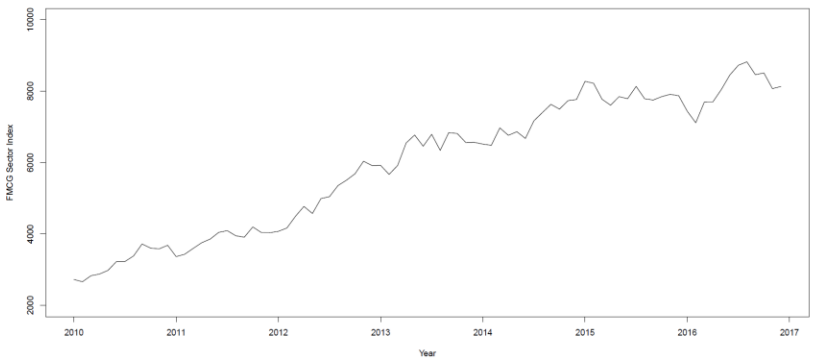


**Figure1.** *Time series of FMCG Sector Index in India (Period: Jan 2010 – Dec 2016)*

To obtain further insights into the characteristics of the time series, we decomposed the time series object into its three components – trend, seasonal and random. The decomposition of the time series object is done using the *decompose( )* function defined in the TTR library in R programming environment. The *decompose( )* function is executed with the FMCG time series object as its parameter

and the three components of the time series are obtained. Figure 2 presents the graphs of FMCG sector time series and its three components. Figure 2 consisted of four boxes arranged in a stack. The boxes display the overall time series, the trend, the seasonal and the random component respectively arranged from top to bottom in that order.

From Figure 1, it may be seen that the time series of the FMCG sector consistently increased during the period January 2010 till December 2014. During the year 2015, the time series was rather flat in nature. However, during the period January 2016 till August 2016, the FMCG sector time series exhibited a positive slope before it started experiencing a fall that continued till the end of the year 2016. Figure 2 shows the decomposition results of the FMCG time series. The three components of the time series are shown separately so that their relative behavior can be visualized.

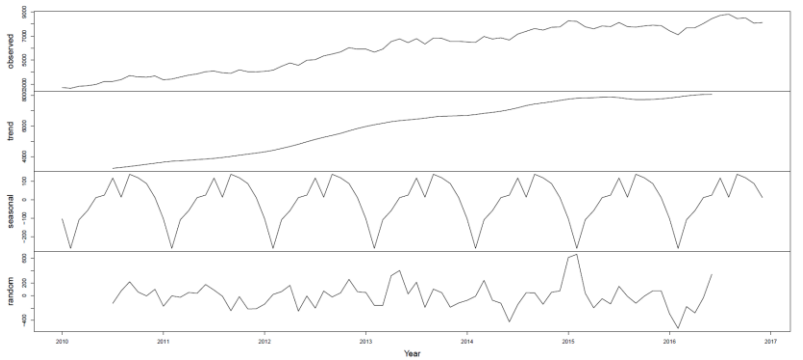


**Figure 2.** *Decomposition results of the FMCG sector index time series into its three components (Period: Jan 2010 – Dec 2016)*

Table 1 presents the numerical values of the time series data and its three components. The trend and the random components are not available for the period January 2010 – June 2010 and also for the period July 2016 – December 2016. This is due to the fact that trend computation needs long

term data. Coughlan (2015) illustrated that the *decompose( )* function in R uses a 12-month moving average method to compute the trend values in a time series. Hence, in order to compute trend values for the period January2010 – June 2010, the *decompose( )* function needs time series data for the period July 2009 – December 2009. Since, the data for the period July 2009 – December 2009 are not available in the dataset under our study, the trend values for the period January 2010 – June 2010 could not be computed. For similar reason, the trend values for the period July 2016 - December 2016, could not be computed due to non- availability of the time series record for the period January 2017 – June 2017 in our dataset. It may be noted from Table 1 that the seasonal value for a given month remains constant throughout the entire period of study. For example, the seasonal component has a constant value of 22 for the month of January in every year from 2010 till 2016. It is interesting to note that due to the non-availability of the trend values for the periods January 2010 – June 2010 and July 2016 – December 2016, the random components for these periods could not also be computed by the *decompose( )* function. In other words, since the aggregate time series values are given by the sum of the corresponding trend, seasonal and random component values, and because of the fact that the seasonal value for a given month remains the same throughout, non-availability of trend values for a period makes the random components values also unavailable for the same period.

**Table 1.** *Time series index values of the Indian FMCG Sector and its components (Period: Jan 2010 – Dec 2016)*

| Year | Month | Aggregate Index | Trend | Seasonal | Random |
|---|---|---|---|---|---|
| 2010 | January | 2725 | | -103 | |
| | February | 2662 | | -263 | |
| | March | 2831 | | -108 | |
| | April | 2878 | | -58 | |
| | May | 2981 | | 12 | |
| | June | 3230 | | 26 | |

| | | | | | |
|---|---|---|---|---|---|
| | July | 3230 | 3236 | 119 | -125 |
| | August | 3385 | 3295 | 15 | 75 |
| | September | 3720 | 3359 | 139 | 222 |
| | October | 3605 | 3427 | 120 | 57 |
| | November | 3583 | 3500 | 89 | -6 |
| | December | 3684 | 3571 | 13 | 101 |
| 2011 | January | 3366 | 3641 | -103 | -172 |
| | February | 3432 | 3700 | -263 | -5 |
| | March | 3596 | 3732 | -108 | -28 |
| | April | 3755 | 3764 | -58 | 49 |
| | May | 3858 | 3808 | 12 | 37 |
| | June | 4045 | 3842 | 26 | 177 |
| | July | 4093 | 3886 | 119 | 88 |
| | August | 3950 | 3946 | 15 | -11 |
| | September | 3910 | 4014 | 139 | -243 |
| | October | 4197 | 4094 | 120 | -17 |
| | November | 4041 | 4166 | 89 | -214 |
| | December | 4035 | 4235 | 13 | -213 |
| 2012 | January | 4074 | 4315 | -103 | -137 |
| | February | 4167 | 4413 | -263 | 17 |
| | March | 4493 | 4538 | -108 | 63 |
| | April | 4772 | 4667 | -58 | 164 |
| | May | 4574 | 4812 | 12 | -250 |
| | June | 4992 | 4973 | 26 | -7 |
| | July | 5046 | 5129 | 119 | -202 |
| | August | 5356 | 5268 | 15 | 73 |
| | September | 5507 | 5390 | 139 | -23 |
| | October | 5687 | 5524 | 120 | 43 |
| | November | 6038 | 5689 | 89 | 260 |
| | December | 5916 | 5842 | 13 | 61 |
| 2013 | January | 5922 | 5976 | -103 | 49 |
| | February | 5669 | 6090 | -263 | -158 |
| | March | 5919 | 6186 | -108 | -159 |
| | April | 6549 | 6289 | -58 | 319 |
| | May | 6772 | 6358 | 12 | 402 |
| | June | 6458 | 6407 | 26 | 26 |
| | July | 6792 | 6458 | 119 | 215 |
| | August | 6342 | 6517 | 15 | -190 |
| | September | 6838 | 6595 | 139 | 104 |
| | October | 6814 | 6648 | 120 | 46 |
| | November | 6562 | 6661 | 89 | -187 |
| | December | 6567 | 6673 | 13 | -119 |
| 2014 | January | 6518 | 6698 | -103 | -77 |
| | February | 6484 | 6758 | -263 | -11 |
| | March | 6971 | 6835 | -108 | 244 |
| | April | 6763 | 6897 | -58 | -75 |
| | May | 6864 | 6974 | 12 | -123 |
| | June | 6676 | 7073 | 26 | -423 |

| | | | | | |
|---|---|---|---|---|---|
| | July | 7170 | 7196 | 119 | -145 |
| | August | 7402 | 7342 | 15 | 45 |
| | September | 7631 | 7448 | 139 | 44 |
| | October | 7497 | 7516 | 120 | -140 |
| | November | 7734 | 7592 | 89 | 53 |
| | December | 7767 | 7680 | 13 | 75 |
| 2015 | January | 8275 | 7766 | -103 | 612 |
| | February | 8222 | 7823 | -263 | 663 |
| | March | 7773 | 7844 | -108 | 37 |
| | April | 7607 | 7863 | -58 | -198 |
| | May | 7847 | 7885 | 12 | -51 |
| | June | 7789 | 7897 | 26 | -134 |
| | July | 8134 | 7867 | 119 | 149 |
| | August | 7788 | 7786 | 15 | -12 |
| | September | 7752 | 7736 | 139 | -123 |
| | October | 7847 | 7736 | 120 | -10 |
| | November | 7912 | 7748 | 89 | 75 |
| | December | 7872 | 7784 | 13 | 75 |
| 2016 | January | 7439 | 7837 | -103 | -295 |
| | February | 7114 | 7904 | -263 | -527 |
| | March | 7692 | 7977 | -108 | -177 |
| | April | 7697 | 8034 | -58 | -279 |
| | May | 8045 | 8069 | 12 | -36 |
| | June | 8453 | 8086 | 26 | 341 |
| | July | 8725 | | 119 | |
| | August | 8822 | | 15 | |
| | September | 8461 | | 139 | |
| | October | 8511 | | 120 | |
| | November | 8071 | | 89 | |
| | December | 8131 | | 13 | |

From Table 1, some important observations are made. First, we can see that seasonal component has the maximum value of 139 in the month of September, while the lowest value of seasonality -263 is observed in the month of February. The seasonal component is found to have high positive values during the months of July, September and October, while negative values of seasonality are observed during the months of January till April. In order to analyze the impact of seasonality on the FMCG time series, we computed some statistics on the seasonal component values. We computed the percentage contribution of the seasonal component on the aggregate time series values and found

the following. The maximum, the minimum, the mean of the absolute values, and the standard deviation of the percentage of the seasonal components with respect to the aggregate time series values were found to be 3.74, -9.88, 1.72 and 2.42 respectively. The maximum percentage of seasonal component was found in the month of September 2010, while the minimum was observed in the month of February 2010. While the low value of the mean percentage indicated that the time series was not seasonal, the high value of standard deviation in comparison to the mean value implied that the seasonal percentages exhibited high level of dispersion among themselves.

Second, the trend component of the time series consistently increased during the period July 2010 till December 2014. However, the trend started became sluggish and flattened out from January 2015 and continued to maintain the same pattern till July 2016. The maximum, the minimum, the mean, and the standard deviation of the percentage of the trend component with respect to the aggregate time series were found to be 111.10, 90.30, 100.16 and 3.94 respectively indicating that the trend was the single most dominant component in the time series. The maximum percentage of trend component was found in the month of February 2016, while the minimum was found in the month of February 2010.

Third, the maximum and the minimum values of the random component of the time series were found to be 663 and -527 respectively. These values are quite modest in comparison to the aggregate time series values. In order to understand the contribution of the random component on the overall time series, we computed the maximum, the minimum, the mean of the absolute values and the standard deviation of the percentage of random component values with respect to the aggregate time series values. These values were found to be 8.06, -7.41, 2.42 and 3.19 respectively. It

indicated that while the random component is not dominant in the time series, the values of the random component exhibited large deviations across their mean value. The random component contributed its maximum percentage to the aggregate timeseries in the month of December 2014, while the lowest percentage was found in the month of February 2016.

The overall conclusion is that the FMCG time series is primarily dominated by its trend component, while seasonal and random components are having not significant contributions to the aggregate time series. However, the seasonal and random components exhibited significant variations across their mean values.

## Proposed forecasting methods

In this Section, we present a set of interesting forecasting methods that we have applied on the time series data of the FMCG sector index. We propose six different approaches to forecasting and present the performance of these approaches on the FMCG sector time series data. For the purpose of comparative analysis of different approaches of forecasting, we use five different metrics and identify which method leads to the lowest value of forecasting error. We also critically analyze the approaches and argue why one method performs better than the others on the given dataset of FMCG sector time series index for the period January 2010 – December 2016. In the following, we first describe the six approaches, and then provide the detailed results as these forecasting methods are applied on the FMCG sector dataset.

**Method 1:** In this method, we use the FMCG sector time series data for the period January 2010 till December 2015 for the purpose of forecasting the monthly index values for each month of the year 2016. The *HoltWinters( )* function in R library *forecast*has been used for this purpose. In order to build a robust forecasting framework, the *HoltWinters* model

is used with a *changing trend* and an *additive seasonal* component that best fits the FMCG time series index data. The forecast *horizon* in the *HoltWinters* model is chosen to be 12 so that the forecasted values for all months of 2016 can be obtained by using the method at the end of the year 2015. Forecast error is computed for each month of 2016 and an overall RMSE value is also derived for this method.

**Method II:** In this approach, the FMCG sector index value for each month of the year 2016 is forecasted using the *HoltWinters( )* method with a forecast horizon of 1 month. For example, for the purpose of forecasting the index for the month of March 2016, the index values of the FMCG sector from January 2010 till February 2016 are used to develop the forecasting model. As in Method I, the *HoltWinters* model is used with a changing trend and an additive seasonal component. Since the forecast horizon is short, the model is likely to produce higher accuracy in forecasting compared to the approach followed in Method I that used a forecast horizon of 12 months. The forecast error corresponding to each month of 2016 and an overall RMSE value for the model is computed.

**Method III:** In this forecasting approach, we first use the time series data for the FMCG sector index for the period January 2010 till December 2015 and derive the trend and the seasonal component values. This method yields the values of the trend component of the time series for the period from July 2010 till June 2015. Using the series of trend values for the period July 2010 till June 2015, forecast for the trend values for the period July 2015 till June 2016 are made using the *HoltWinters( )* function in R with a changing trend component level but without any seasonality component. In other words, for forecasting the trend values, in the *HoltWinters( )* function in R, we set the parameter *'beta'* = TRUE and the parameter '*gamma*' = FALSE, in the *HoltWinters( )* function in R. The forecasted trend values are

added to the seasonal component values of the corresponding months (based on the time series data for the period January 2010 till December 2015) to arrive at the forecasted aggregate of the trend and seasonal components. Now, we consider the time series of the FMCG sector index values for the entire period, i.e., from January 2010 till December 2016, and decompose it into its trend, seasonal and random components. Based on thistime series, we compute the aggregate of the actual trend and the actual seasonal component values for the period July 2015 till June 2016. We derive the forecasting accuracy of this method by calculating the percentage of deviation of the aggregate of the actual trend and the actual seasonal component values with respect to the corresponding aggregate values of the forecasted trend and past seasonal components for each month during July 2015 to June 2016. An overall RMSE value for this method is also computed.

**Method IV:** The approach followed in this method is exactly similar to that used in Method III. However, unlike Method III that used *HoltWinters( )* function with changing trend component and a zero seasonal component, this method uses a linear regression model for the purpose of forecasting the trend component values for the period from July 2015 till June 2016. The *lm ( )* function in R is used for building a bivariate linear regression model with trend component as the response variable and time as the predictor variable. The regression model is built using the trend values for the period July 2010 till June 2015. The aggregate of the predicted trend values and the past seasonal values are compared with the aggregate of the actual trend values and the actual seasonal values for the period from July 2015 till June 106. The error in forecasting and an overall RMSE value is computed as in Method III.

**Method V:** We use *Auto Regressive Integrated Moving Average* (ARIMA) based approach of forecasting in this

method. For building the ARIMA model, we use the FMCG sector time series data for the period January 2010 till December 2015. Using this training data set and executing the *auto.arima( )* function defined in the forecast package in R, we compute values of the three parameters of the *Auto Regressive Moving Average* (ARMA) model, i.e. the *autoregression* parameter (*p*), the *difference* parameter (*d*), and the *moving average* parameter (*q*). Next, we build the ARIMA model of forecasting using the *arima( )* function in R with the two parameters: (i) the FMCG time series R object (based on data for the period from January 2010 till December 2015), (ii) the order of the ARMA i.e., the values of the three parameters (*p*, *d*, *q*).Using the resultant ARIMA model, we call the function *forecast.Arima( )* with parameters: (i) ARIMA model object, and (ii) forecast horizon = 12 months (in this approach). Since a forecast horizon of 12 months is used, we compute the forecasted values of each month of the year 2016 at the end of the year 2015. The error in forecasting and the RMSE value are also computed.

**Method VI:** Similar to Method V, this forecasting method is also based on an ARIMA model. However, unlike Method V that used forecast horizon of 12 months, this method uses a short forecast horizon of 1 month. For the purpose of forecasting, the ARIMA model is built using time series data for the period January 2010 till the month previous to the month for which forecasting is being made. For example, for the purpose of prediction of the time series index for the month of May 2016, the time series data from January 2010 till April 2016 is used for building the ARIMA model. Since the training data set for building the ARIMA model constantly changes in this approach, we evaluate the ARIMA parameters (i.e., *p*, *d*, and *q*) before every round of forecasting. In other words, for each month of the year 2016, before we make the forecast for the next month, we compute the values of the three parameters of the ARIMA model.

## Forecasting results

In this Section, we provide results on the performance of the six forecasting methods.

**Method I:** The results obtained using these methods are presented in Table 2. In Figure 3, we have plotted the actual values of the FMCG sector index and the corresponding predicted values for each month of the year 2016.

**Table 2.** *Computation Results using Method I*

| Month | Actual Index | Forecasted Index | % Error | RMSE |
|---|---|---|---|---|
| (A) | (B) | (C) | (C-B)/B *100 | |
| Jan | 7439 | 7928 | 6.57 | |
| Feb | 7114 | 7727 | 8.62 | |
| Mar | 7692 | 7767 | 0.98 | |
| Apr | 7697 | 7915 | 2.83 | |
| May | 8045 | 8126 | 1.01 | |
| Jun | 8453 | 8110 | 4.06 | 373 |
| Jul | 8725 | 8415 | 3.55 | |
| Aug | 8822 | 8259 | 6.38 | |
| Sep | 8461 | 8469 | 0.09 | |
| Oct | 8511 | 8550 | 0.46 | |
| Nov | 8071 | 8601 | 6.57 | |
| Dec | 8131 | 8556 | 5.23 | |

**Observations on Method I:** We observe from Table 2 that the forecasted values very closely match the actual values of the FMCG sector index for all months in the year 2016. The highest value of the error percentage (8.62) has been found to be quite low which corresponds to the month of February 2016. The forecast for the month of September 2016 exhibited the lowest percentage of error (0.09), which is almost negligible. The reason for such accuracy for this method of forecasting is due to the modest growth of the FMCG sector during the year 2016. Since there was no sharp rise or fall of the actual index value in any month of 2016, the *HoltWinters* method of forecasting with a long forecast horizon of 12 months that effectively forecasted a smoothened average

value turned out to be a very effective method. The RMSE value for this method is found out to be 373 that is only 4.61 percent of the mean value of 8097 of the actual index of the FMCG sector index during 2016. This indicates that Method I is very accurate in forecasting the FMCG sector time series index.

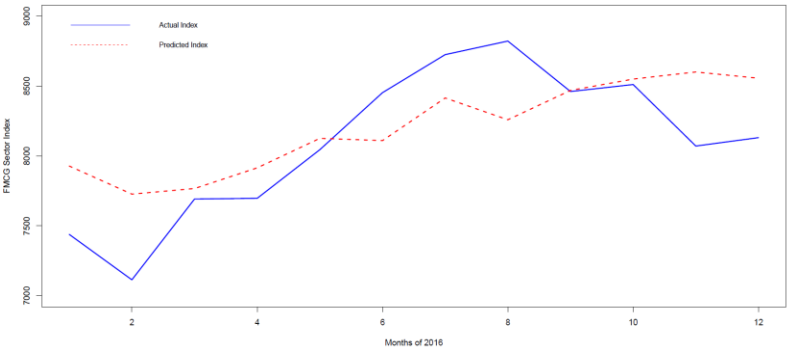


**Figure 3.** *Actual and predicted values of FMCG sector index using Method I of forecasting. (Period: Jan 2016 – Dec 2016).*

**Method II:** The results of forecasting using Method II are presented in Table 3. In Figure 4, the actual index values and their corresponding predicted values are plotted.

**Table 3.** *Computation Results using Method II*

| Month | Actual Index | Forecasted Index | % Error | RMSE |
|---|---|---|---|---|
| (A) | (B) | (C) | (C-B)/B *100 | |
| Jan | 7439 | 7928 | 6.57 | |
| Feb | 7114 | 7395 | 3.95 | |
| Mar | 7692 | 7248 | 5.77 | |
| Apr | 7697 | 7723 | 0.34 | |
| May | 8045 | 7883 | 2.01 | |
| Jun | 8453 | 8024 | 5.08 | 343 |
| Jul | 8725 | 8555 | 1.95 | |
| Aug | 8822 | 8558 | 2.99 | |
| Sep | 8461 | 9000 | 6.37 | |
| Oct | 8511 | 8682 | 2.01 | |
| Nov | 8071 | 8588 | 6.41 | |
| Dec | 8131 | 8165 | 0.42 | |

**Observations on Method II:** We observe from Table 3 and Figure 4 that the forecasted values very closely match with the actual values of the time series index. The lowest value of error percentage is found to be 0.34 which occurred in the month of April 2016, while the highest error percentage value of 6.57 was found in the month of January 2016. The RMSE value for this method is 343, which even lower than that obtained in Method I earlier. The RMSE value contributes only 4.24 percent of the mean value of the actual index of the FMCG sector for the year 2016. This clearly demonstrates that *HoltWinters* additive model with a prediction horizon of 1 month can very effectively and accurately forecast FMCG sector index values.

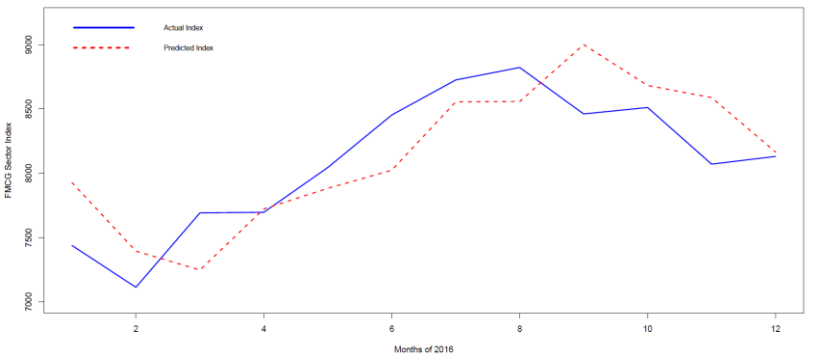


**Figure 4.** *Actual and predicted values of FMCG sector index using Method II of forecasting.(Period: Jan 2016 – Dec 2016).*

**Table 4.** *Computation Results using Method III*

| Month | Actual Trend | Actual Seasonal | Actual (Trend + Seasonal) | Fore-casted Trend | Past Seasonal | Forecasted (Trend + Seasonal) | % Error | RMSE |
|---|---|---|---|---|---|---|---|---|
| A | B | C | D | E | F | G | (G-D)/D *100 | |
| Jul | 7867 | 119 | 7986 | 7909 | 75 | 7984 | 0.02 | |
| Aug | 7786 | 15 | 7801 | 7921 | 4 | 7925 | 1.59 | |
| Sep | 7736 | 139 | 7875 | 7933 | 150 | 8083 | 2.64 | 132 |
| Oct | 7736 | 120 | 7856 | 7945 | 109 | 8054 | 2.52 | |
| Nov | 7748 | 89 | 7837 | 7957 | 60 | 8017 | 2.30 | |
| Dec | 7784 | 13 | 7797 | 7969 | -16 | 7953 | 2.00 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Jan | 7837 | 103 | 7940 | 7981 | -58 | 7923 | 0.21 |
| Feb | 7904 | -263 | 7641 | 7993 | -172 | 7821 | 2.36 |
| Mar | 7977 | -108 | 7869 | 8005 | -86 | 7919 | 0.64 |
| Apr | 8034 | -58 | 7976 | 8017 | -16 | 8001 | 0.31 |
| May | 8069 | 12 | 8081 | 8029 | 6 | 8035 | 0.57 |
| Jun | 8086 | 26 | 8112 | 8041 | -56 | 7985 | 1.57 |

**Method III:** The results of forecasting using this method are presented in Table 4. Figure 5 depicts the actual index values and their corresponding predicted values for all months of the year 2016 using Method III of forecasting. The actual trend and seasonal component values for the period from July 2015 till June 2016 (computed based on the time series data for the period from January 2010 till December 2016) and their aggregated monthly values are noted in Columns *B*, *C* and *D* respectively in Table 4. The forecasted trend values (using *HoltWinters* method with changing trend component and nil seasonal component and with a forecast horizon of 12 months) and the past seasonal component values (based on time series data for the period from January 2010 till December 2015) and their corresponding aggregate values are noted in columns *E*, *F* and *G* respectively in Table 4. The error value for each month and an overall RMSE value for this method are also computed.

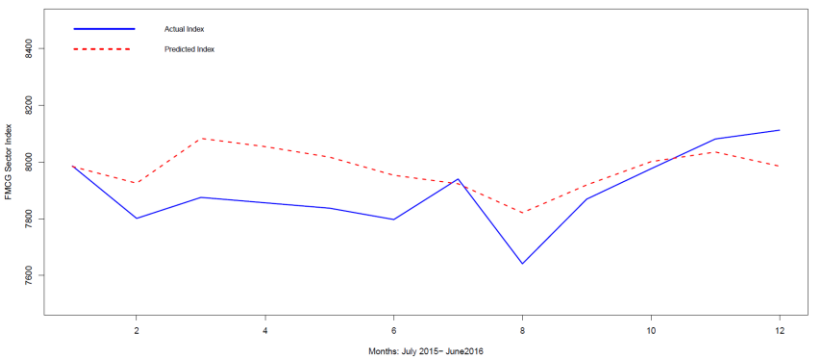


**Figure 5.** *Actual and predicted values of the sum of trend and seasonal components of FMCG sector index using Method III of forecasting (Period: Jan 2016 – Dec 2016)*

**Observations on Method III:** It may be observed from Table 4 and Figure 5 that the error in forecasting has been quite small for all months during the period July 2015 till June 2016. The lowest value of the error percentage had been 0.02 that occurred in the month of July 2015, while the highest value of error percentage of 2.64 was observed in the month of September 2015. The computed RMSE value of 132 for this method represents only 1.67 percent of the mean value of the sum of the actual trend and the actual seasonal components during the period July 2015 till June 2016. The low value of RMSE indicates that this method is very effective in forecasting the FMCG index time series index values. It is evident from Table 4 that the actual trend values of the FMCG sector increased very sluggishly over the period July 2015 till June 2016. Since the actual seasonal values are quite small compared to the trend values, the sum of the actual trend and the actual seasonal values also exhibited the same pattern of change as the trend, increasing at a very slow rate starting with a value of 7986 in July 2015 attaining a value of 8112 in June 2016. As the trend values are forecasted using *HoltWinters* method with a forecast horizon of 12 months, the forecasted trend values are smoothened out over the period, resulting in a series of forecasted trend values exhibiting a very slow rate of increase. In fact, the forecasted trend started with a value of 7909 in the month of July 2015, attained a value of 8041 in June 2016. Since the past seasonal values (i.e., the seasonal values based on the FMCG sector time series for the period January 2010 till December 2015), were also very low compared to the corresponding trend values, the sum of the forecasted trend values and their respective past seasonal values increased very sluggishly exhibiting the same pattern as the forecasted trend values. The slow rate of increase for both the actual series and the forecasted series made the two series match closely with each other, thereby making this

method of forecast extremely accurate for the FMCG sector index time series.

**Method IV:** Table 5 presents the results of forecasting for Method IV. Figure 6 shows the actual index values and their corresponding predicted values using Method IV of forecasting for all months during the period July 2015 till June 2016. The values of the actual trend component and the actual seasonal component for the period July 2015 till June 2016 (computed based on the time series data for the period January 2010 till December 2016) and their aggregated monthly values are listed in Columns *B*, *C* and *D* respectively in Table 4. In Method III, the trend values are forecasted using *HoltWinters( )* function with a forecast horizon of 12 months. However, in Method IV, the forecasting of the trend values are done using a linear regression approach with the FMCG *index* values as the response variable and the *month* as the predictor variable. The forecasted trend values and the past seasonal component values (based on time series data for the period from January 2010 till December 2015) and their corresponding aggregate values listed noted in columns *E*, *F* and *G* respectively in Table 5. The percentage of error in forecasting for each month during the period July 2015 till June 2016, and an overall RMSE value are also listed.

**Table 5.** *Computation Results using Method IV*

| Month | Actual Trend | Actual Seasonal | Actual (Trend + Seasonal) | Fore-caste Trend | Past Seasonal | Forecaste (Trend + Seasonal) | % Error | RMSE |
|---|---|---|---|---|---|---|---|---|
| A | B | C | D | E | F | G | (G-D)/D *10( | |
| Jul | 7867 | 119 | 7986 | 8299 | 75 | 8374 | 4.86 | |
| Aug | 7786 | 15 | 7801 | 8386 | 4 | 8390 | 7.55 | |
| Sep | 7736 | 139 | 7875 | 8474 | 150 | 8624 | 9.51 | |
| Oct | 7736 | 120 | 7856 | 8562 | 109 | 8671 | 10.37 | 927 |
| Nov | 7748 | 89 | 7837 | 8649 | 60 | 8709 | 11.13 | |
| Dec | 7784 | 13 | 7797 | 8737 | -16 | 8721 | 11.85 | |
| Jan | 7837 | -103 | 7734 | 8824 | -58 | 8766 | 13.34 | |
| Feb | 7904 | -263 | 7641 | 8912 | -172 | 8740 | 14.38 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Mar | 7977 | -108 | 7869 | 8999 | -86 | 8913 | 13.27 |
| Apr | 8034 | -58 | 7976 | 9087 | -16 | 9071 | 13.73 |
| May | 8069 | 12 | 8081 | 9174 | 6 | 9180 | 13.60 |
| Jun | 8086 | 26 | 8112 | 9262 | -56 | 9206 | 13.49 |

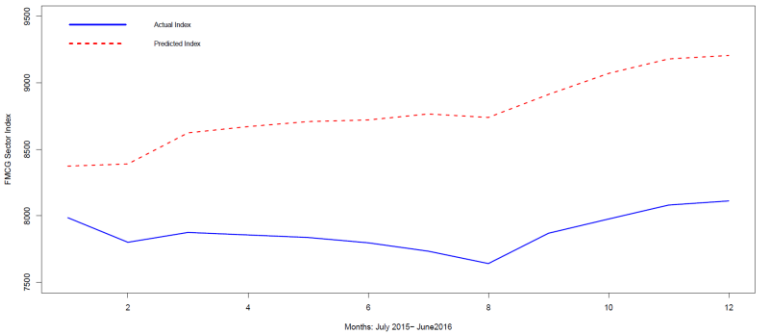


**Figure 6.** *Actual and predicted values of the sum of trend and seasonal components of FMCG sector index using Method IV of forecasting (Period: Jan 2016 – Dec 2016)*

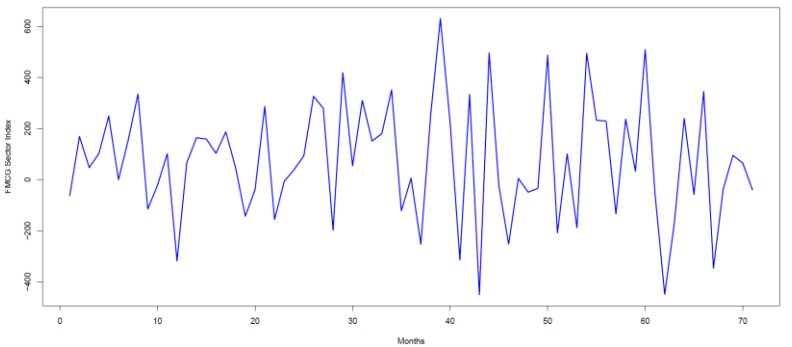


**Figure 7.** *First-order difference of the FMCG sector time series (Period: Jan 2016 – Dec 2016)*

**Observation on Method IV:** It is clear from Table 5 and Figure 6 that unlike Method III, Method IV has produced high values of error. The lowest percentage of error (4.86) was found in the month of July 2015, while the highest error percentage value (13.73) was observed in the month of April 2016. It may also be noted from Figure 5 that the error increased consistently with time. The RMSE value for this

method has been found to be 927 while the mean value of the actual sum of trend and seasonal components during the period July 2015 till June 2016 is 7897. Hence, the RMSE value is 11.74 percent of the mean value of the index, indicating that the method of forecasting is effective although it involves higher error than the methods discussed earlier. It is not difficult to understand the reason for the higher values of error produced by this method. This method used linear regression method for predicting the trend values for the period July 2015 till June 2016. Based on the previous trend values for the period July 2010 till June 2015, the linear regression computed the regression coefficients and used the values of the coefficients to compute the trend values for the period July 2015 till June 2106. Since, the trend for the period July 2010 till June 2015 had a positive slope, the regression coefficients were positive. The positive regression coefficients produced forecasted values of trend which consistently increased from a value of 8299 in July 2015 to 9262 in June 2016. Since seasonal component values were small compared to trend values, the sum of the forecasted trend and past seasonal values also exhibited the same pattern as the forecasted trend values – the sum of the forecasted trend values and the past seasonal values increased consistently from 8374 in July 2015 to 9206 in June 2016. However, the actual trend values during the target period of July 2015 till June 2016 were very sluggish. The actual trend increased very slowly from a value of 7867 in July 2015 to 8086 in July 2016. Since the forecasted trend values increased a much faster rate than the actual rate of increase of the trend time series, the method yielded increasingly higher values of error with time.

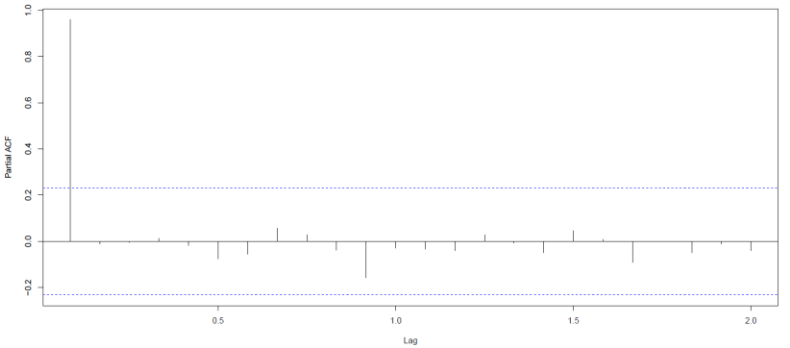


**Figure 8.** *Plot of the partial auto correlation function (PACF) of the FMCG sector time series with max lag of 2 years (Period: Jan 2016 – Dec 2016)*

**Method V:** This method of forecasting is based on ARIMA technique with a forecast horizon of 12 months. Applying *auto.arima( )* function defined in the library *forecast* in R, on the FMCG sector time series index values for the period January 2010 till December 2015, we obtain the parameter values for the time series as: $p = 0$, $d = 1$, and $q = 1$. We cross verify the values of $p$, $d$, and $q$ by plotting the *partial auto correlation function* (PACF), the *first-order difference* of the time series, and the *auto correlation function* (ACF) respectively. The first order difference of the FMCG time series is presented in Figure 7. It is clear that the first-order difference time series is a stationary one, as the mean and the variance of the first-order difference time series are approximately constant. Hence the value of $d = 1$ is cross-verified. Now, we plot the *partial auto correlation function* (PACF) and the *auto correlation function* (ACF) to cross-check the values of the parameters $p$ and $q$. Figure 8 depicts the PACF of the FMCG sector time series. It is clear that except for *lag* = 0, partial correlation values at all lags are insignificant. Hence the value of $p = 0$ is also verified. Figure 9 shows that minimum integral value of lag beyond which all autocorrelation values are insignificant is 1. Therefore, $q$ =

1 is also verified. Hence, it now confirmed that the FMCG sector time series for the period January 2010 till December 2015 can be modeled as an ARMA (0, 1, 1) model. Using the *arima( )* function in R with its two parameters: (i) the FMCG sector time series R object and (ii) the order (0, 1, 1) of ARMA, we build the ARIMA model. Finally, we use the function *forecast. Arima( )* with two parameters: (i) the ARIMA model and (ii) the time horizon of forecast = 12 months, for forecasting the index values of the time series for all the twelve months of the year 2016. Table 6 presents the results of forecasting using this method, while Figure 10 depicts the actual index values and their corresponding predicted values for all months in the year 2016.

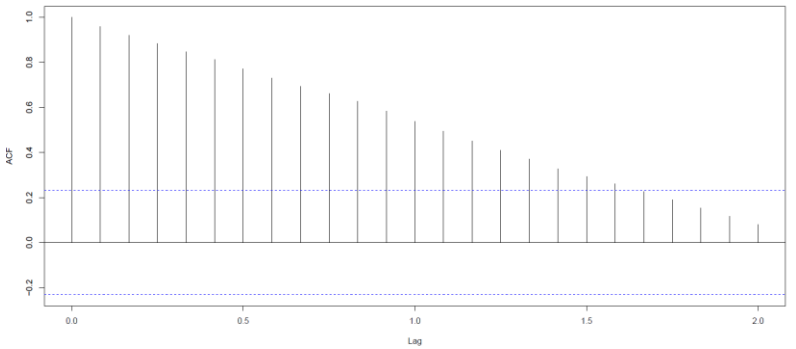


**Figure 9.** *Plot of the auto correlation function (ACF) of the FMCG sector time series with max lag of 2 years (Period: Jan 2016 – Dec 2016)*

**Table 6.** *Computation Results using MethodV*

| Month | Actual Index | Forecasted Index | % Error | RMSE |
|---|---|---|---|---|
| (A) | (B) | (C) | (C-B)/B *100 | |
| Jan | 7439 | 7875 | 5.86 | |
| Feb | 7114 | 7875 | 10.70 | |
| Mar | 7692 | 7875 | 2.38 | |
| Apr | 7697 | 7875 | 2.31 | |
| May | 8045 | 7875 | 2.11 | 553 |
| Jun | 8453 | 7875 | 6.84 | |
| Jul | 8725 | 7875 | 9.74 | |
| Aug | 8822 | 7875 | 10.73 | |
| Sep | 8461 | 7875 | 6.93 | |
| Oct | 8511 | 7875 | 7.47 | |

| | | | |
|---|---|---|---|
| Nov | 8071 | 7875 | 2.43 |
| Dec | 8131 | 7875 | 3.15 |

**Observations:** It is evident from Table 6 that most of the error percentage values are quite moderate. The lowest value of error percentage had been 2.11 that occurred in the month of May 2016, while the highest value of error percentage was 10.73, observed in the month of August 2016. The RMSE value for this method is found to be 553 which is 6.83 per cent of the mean value of the FMCG sector index during the period January 2016 till December 2016. The mean value of the FMCG sector index has been 8097. Considering the fact that this method uses a long forecast horizon of 12 months, the error percentage values are quite low. This is attributed to the fact that the FMCG sector index experienced a very modest increase during the period January 2016 till December 2016 – the index started with a value of 7439 in January 2016 and attained a value of 8131 in December 2016. The *HoltWinters* method with forecast horizon of 12 forecasted a constant average value of 7875 for the series with ARIMA parameters (0, 1, 1) so that the average error for the entire series is minimized. Figure 10 presents a graphical depiction of the actual index values and their corresponding predicted values using Method V.

**Method VI:** In this approach, we build an ARIMA model with a forecast horizon of one month. The methodology used for building the ARIMA model, however, is exactly identical to that used in Method V. The difference in Method V and Method VI lies in different values of forecast horizon used in these methods. While Method V uses a forecast horizon of 12 months, we use a forecast horizon of 1 month in Method VI. Since, in Method VI, forecast is made only one month in advance, the training data set used for building the ARIMA model constantly increases in size, and hence, we re-evaluate the parameters of the ARIMA model every time we use it in

forecasting. In other words, for each month of 2015, before we make the forecast for the next month, we compute the values of the parameters of the ARIMA model. Computation of the values of ARIMA parameters *p*, *d*, and *q* showed that for the period January 2016 till June 2016, the ARIMA model was (0, 1, 1), while the for the remaining period July 2016 till December 2016, it was (0, 1, 0). Table 7 presents the forecasting results for Method VI. Figure 11 depicts the actual index values of the FMCG sector and their corresponding predicted values for this method of forecasting.

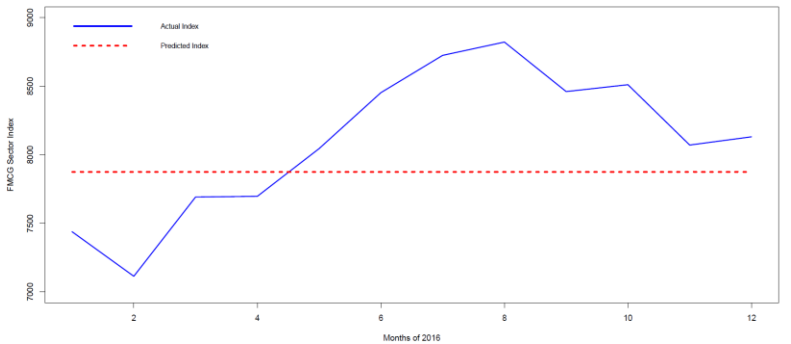


**Figure 10**. *Actual and predicted values of FMCG sector index using Method V of forecasting. (Period: Jan 2016 – Dec 2016).*

**Table 7.** *Computation Results using Method VI*

| Month | Actual Index | Forecasted Index | % Error | RMSE |
|---|---|---|---|---|
| (A) | (B) | (C) | (C-B)/B *100 | |
| Jan | 7439 | 7875 | 5.86 | |
| Feb | 7114 | 7475 | 5.07 | |
| Mar | 7692 | 7114 | 7.51 | |
| Apr | 7697 | 7640 | 0.74 | |
| May | 8045 | 7692 | 4.39 | |
| Jun | 8453 | 8016 | 5.17 | 340 |
| Jul | 8725 | 8453 | 3.12 | |
| Aug | 8822 | 8725 | 1.10 | |
| Sep | 8461 | 8822 | 4.27 | |
| Oct | 8511 | 8461 | 0.59 | |
| Nov | 8071 | 8511 | 5.45 | |
| Dec | 8131 | 8071 | 0.74 | |

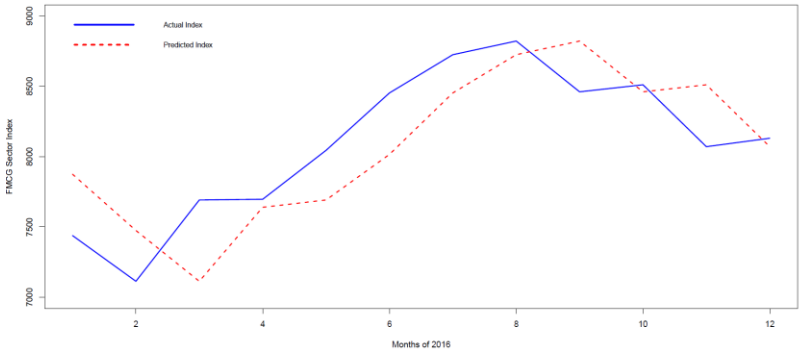


**Figure 11.** *Actual and predicted values of FMCG sector index using Method VI of forecasting. (Period: Jan 2016 – Dec 2016).*

**Observations:** From Table 7, it is evident that the error percentage values for all months of the year 2016 are very low. The lowest value of error percentage was found to be 0.59 in the month of October 2016, while the highest value of error was 7.51 per cent in the month of March 2016. The RMSE value for Method VI of forecasting is found to be 340. The mean value of the index of the FMCG sector for the period January 2016 till December 2016 is 8097. Hence, the RMSE value is just 4.20 percent of the mean value of the actual index of the FMCG sector. This indicates that Method VI has been highly accurate in forecasting the FMCG sector index values. The high level of accuracy of Method VI may be attributed to its short forecast horizon of one month. The short forecast horizon is able to catch the changing pattern of the time series very effectively. This has resulted into a very small error in forecasting. It is clearly evident from Figure 11 that the forecasted time series values exactly followed the pattern of the time series of the actual index values of the FMCG sector.

## Summary of forecasting results

In Table 8, we summarize the performance of the six forecasting methods that we have used. For the purpose of

comparison between these methods, we have chosen six metrics: (i) minimum (Min) error rate, (ii) maximum (Max) error rate, (iii) mean error rate, (iv) standard deviation (SD) of error rates, and (v) root mean square error (RMSE), and (vi) the ratio of RMSE to the mean of the actual index values in percentage. For Method I, II, IV and V the mean of the index values are the same, being the mean of the actual index values of the FMCG sector for the twelve months in the year 2016. This mean value is found to be equal to 8097. For Method III and IV, however, the mean of the index values are the mean of the sum of the actual trend and seasonal values during the period July 2015 till June 2016. This mean value of the sum of the trend and seasonal values is found to be equal to 7880. Since the RMSE values for Methods I, II, IV and V and those for Method III and IV are computed against different set of actual values, hence, instead of the raw RMSE value, the percentage of RMSE to the mean value of the index serves as a better metric for comparing different methods of forecasting. Accordingly, we have ranked the forecasting methods

Table 8 presents the comparative analysis of the six forecasting methods. It can be seen that Method III that uses the sum of the forecasted trend values using *HoltWinters( )* function of horizon 12 months and the past seasonal values to predict the sum of the future trends values and the new seasonal values, has performed has produced the lowest percentage value of the ratio of RMSE to the mean index value. In fact, Method III has produced lowest values for all other metrics too. Hence, Method III clearly turns out to be the most accurate among all the six methods. On the other hand, the performance of Method IV has been the worst since it has produced the highest values for four metrics: ratio of RMSE to the mean value, min errors, max error and mean error. The performance of Methods I, II and VI are comparable since in terms of all the five metrics, these three

methods are very close to each other. Method V, however, performed worse than these three methods producing a higher value of 6.83 per cent for the ratio of the RMSE to the mean index value.

**Table 8.** *Comparison of the Performance of the Forecasting Methods*

| Metrics / Methods | Min Error | Max Error | Mean Error | SD of Errors | RMSE | RMSE / Mean Index Value Percentage |
|---|---|---|---|---|---|---|
| Method 1 | 0.09 | 8.62 | 3.86 | 2.84 | 373 | 4.61 |
| Method II | 0.34 | 6.57 | 3.66 | 2.34 | 343 | 4.24 |
| Method III | 0.02 | 2.64 | 1.39 | 0.99 | 132 | 1.67 |
| Method IV | 4.86 | 14.38 | 11.42 | 2.91 | 927 | 11.74 |
| Method V | 2.11 | 10.73 | 5.89 | 3.37 | 553 | 6.83 |
| Method VI | 0.59 | 7.51 | 3.67 | 2.36 | 340 | 4.20 |

## Related work

Several approaches and techniques are proposed by researchers in the literature for forecasting of daily stock prices. Among these approaches, neural network-based approaches are extremely popular. Mostafa (2010) proposed a neural network-based technique for predicting movement of stock prices in Kuwait. Kimoto *et al.* (1990) presented a technique using neural network based on historical accounting data and various macroeconomic parameters to forecast variations in stock returns. Leigh *et al.* (2005) demonstrated methods of predicting stock prices and stock market index movements in the New York Stock Exchange (NYSE) during the period 1981 – 1999 using linear regression and simple neural network models. Hammad *et al.* (2009) illustrated how the output of an *artificial neural network* (ANN) model can be made to converge and produce highly accurate forecasting of stock prices. Dutta *et al.* (2006) used ANN models for forecasting closing index values the BSE during the period January 2002 till December 2003. Ying *et al.* (2009) used *Bayesian Network* (BN) – based approach to forecast stock prices of 28 companies listed in DJIA (Dow

Jones Industrial Average) during 1988-1998. Tsai & Wang (2009) illustrated how and why the forecasting accuracy of BN-based approaches usually is higher than that obtained using traditional regression and neural network-based methods. Tseng *et al.* (2012) demonstrated the application of traditional time series decomposition, HoltWintersmodels, Box-Jenkins method and artificial neural networks in forecasting prices of randomly selected 50 stocks during the period September 1998 till December 2010. The study found that forecasting errors values are low for Box-Jenkins method, HoltWinters model and normalized neural network model, while higher values of error were observed in time series decomposition method and non-normalized neural network model. Moshiri & Cameron (2010) built a *back propagation network* (BPN) with econometric models for forecasting the level of inflation using the techniques: (i) Box-Jenkins Autoregressive Integrated Moving Average (BJARIMA) model, (ii) Vector Autoregressive (VAR) model and (ii) Bayesian Vector Autoregressive (BVAR) model. Thenmozhi (2001) applied chaos theory for examining the pattern of changes of stock prices in Bombay Stock Exchange (BSE) during the period August 1980 till September 1997, and found that the daily and weekly returns of BSE index exhibited nonlinear trends, while the movement pattern of the time series of BSE index was found to be weakly chaotic. Hutchinson *et al.* (1994) proposed a novel approach using the principles of *learning networks* for estimating the price of a derivative.

Predictive models based on ANN are found to be extremely accurate in forecasting stock prices. Shen *et al.* (2007), Jaruszewicz & Mandziuk (2004), Ning *et al.* (2009), Pan *et al.* (2005), Hamid & Iqbal (2004), Chen *et al.* (2005), Chen *et al.* (2003), Hanias *et al.* (2007) and de Faria *et al.* (2009) demonstrated the effectiveness of ANN-based models in their forecasting ability of stock price movements. Many

applications of hybrid systems in stock market time series data analysis have also been proposed in the literature. Wu *et al.* (2008), Wang & Nie (2008), Perez-Rodriguez *et al.* (2005), Leung *et al.* (2000) and Kim (2004) proposed applications of hybrid systems in stock price prediction.

In the literature, researchers have also proposed several forecasting techniques which have particularly focused on various issues in the FMCG domain. Lewandowska (2012) presented a work that analyzed a very high level of the shrinkage in the FMCG sector in the European and global markets. The author observed that in order to reduce losses and their impact on the profits of the companies, it is necessary to grant liability for their reduction to the members of the board of directors. Rakicevic & Vujosevic (2015) presented a real-world sales forecasting problem in a FMCG supply chain. The authors used several sales forecasting methods such as: estimation of the last period, average of all observations, moving average, weighted moving average, exponential smoothing, Holt's model of forecasting and Winter's forecasting model. The authors compared these forecasting methods with respect to three metrics: RMSE, *mean absolute percentage error* (MAPE) and *tracking signal* (TS). Vayvay *et al.* (2013) modeled a supply chain for FMCG products and proposed its working styles, tools, organization structure, relations with other departments, procedures and all the processes in management of supply chain. While forecasting the demand of FMCG products, the authors also took into account return of stale or damaged products that are returned from the customer to the company. Several forecasting techniques for supply chain management are also proposed. Vriens & Versteijnen (2007) discussed dynamics of food industry that makes forecasting and planning so critical. The authors also presented various key issues in business planning process such as: planning of promotion for enhanced sales,

collaboration in the upstream supply chain, criticality of the roles of marketing and sales department in demand forecasting of FMCG products, analysis and resolution of capacity bottleneck the supply chain, forecasting of the effectiveness of introduction of new FMCG products etc. Based on these issues, the authors proposed guidelines for implementation of responsive forecasting and planning process in the food industry. Doganis *et al.* (2006) presented a complete framework for developing nonlinear time series sales forecasting model for food products. The framework proposed by the authors is a combination of two artificial intelligence technologies: (i) *radial basis function* (RBF) neural network architecture and (ii) a specially designed *genetic algorithm* (GA). Singhi *et al.* (2015) in their report of the Confederation of Indian Industries (CII) and the Boston Consulting Group (BCG) identified the key trends reshaping the demand, and assessed the impact of these trends on the shape of consumption over the next decade. The authors have also laid down the imperatives for companies going forward to win in the new world of competition. The authors incorporated inputs from leading CEOs on the key trends impacting the FMCG industry, and leveraged proprietary data and research conducted by BCG. Kunc (2005) presented a detailed case study to illustrate competitive dynamics of FMCG industry using a behavioral model.

In contrast to the work mentioned above, our approach in this chapter is based on structural decomposition of time series of the FMCG sector index in India during the period January 2010 till December 2016. Based on the decomposition of the time series, we identified several interesting characteristics of the FMCG sector India. We particularly investigated the nature of the trend, seasonality pattern and degree of randomness of the time series. After analyzing the nature of the FMCG time series, we proposed six forecasting techniques for predicting the index values of

the sector for each month of the year 2016. We computed the accuracies of each of the forecasting techniques, and critically analyzed under what situations a particular technique performs better than the other techniques. Since the forecasting methods proposed in this chapter are all generic in nature, these methods can be very effectively applied in forecasting the future trends and behavior of time series index values of other sectors of economy of India or other countries in the world.

## Conclusion

This chapter has presented a time series decomposition-based approach for analyzing the behavior of the time series of the FMCG sector of the Indian economy during the period January 2010 till December 2016. Algorithms and library-defined functions in the R programming language have been used to decompose the time series index values into three components- trend, seasonal, and random. The decomposition results of the time series provided with several important insights into the behavior exhibited by the FMCG sector time series during the period under our study. Based on the decomposition results, the degree of seasonality and randomness in the time series have been computed. Particularly, it has been possible to identify the months during which the seasonal component in the FMCG time series plays a major role. It is observed that while the month of September experiences the highest seasonality in the FMCG sector, for the month of February the seasonal effect is the lowest. The random component in the time series has been found to be very moderate with the mean value of the percentage of random component to the aggregate time series value being only 2.4. Although, the trend is the most dominant component in the time series, the value of trend increased at a sluggish rate over the period of seven year during January 2010 till December 2016. After a careful

analysis of the decomposition results of the FMCG sector index time series, we have proposed six methods for forecasting the time series index values. The six method of forecasting involved different algorithms of forecasting and different lengths of forecast horizon. It has been observed that Method III that uses the sum of the forecasted trend values using *HoltWinters( )* function of horizon 12 months and the past seasonal values to predict the sum of the future trends values and the new seasonal values has performed best yielding the lowest percentage value of the ratio of RMSE to the mean index value. However, Method IV that predicts the trend values using a linear regression model is found to produce the highest value of the ratio of the RMSE to the mean index value, thereby exhibiting the worst performance among the six methods of forecasting. The performance of HoltWinters method with forecast horizon of 12 months and 1 month and ARIMA method with forecast horizon of 1 month have been quite satisfactory with the RMSE to mean index ratio for these method not exceeding 4.7. However, ARIMA with a forecast horizon of 12 months yielded a higher value of 6.83 percent for the RMSE to mean index value.

While the results in this work provide enough valuable insights into the characteristics of the FMCG index time series in India, and they also serve as guidelines for choosing an appropriate forecasting framework for predicting the future index values of the time series, these results can be extremely useful for constructing an optimized portfolio of stocks. Performing similar exercise on different sectors will enable analysts to understand the individual characteristics of the trend, seasonality and randomness of those sectors. This information can be suitably leveraged by portfolio managers in identifying the timing of buy and sell of stocks from different sectors thereby designing an efficient and optimized portfolio.

## References

Chen, A.-S., Leung, M.T. & Daouk, H. (2003). Application of neural networks to an emerging financial market: forecasting and trading the Taiwan stock index. *Operations Research in Emerging Economics*, 30(6), 901–923. doi. 10.1016/S0305-0548(02)00037-0

Chen, Y., Dong, X. & Zhao, Y. (2005). Stock index modeling using EDA based local linear wavelet neural network. *Proceedings of International Conference on Neural Networks and Brain*, Beijing, China, pp. 1646–1650. doi. 10.1109/ICNNB.2005.1614946

Coghlan, A. (2015). A Little Book of R for Time Series, Release 02. Accessed on: May 10, 2017. [Retrieved from].

de Faria, E.L., Albuquerque, M.P., Gonzalez, J.L., Cavalcante, J.T.P., & Albuquerque, M.P. (2009). Predicting the Brazilian stock market through neural networks and adaptive exponential smoothing methods. *Expert Systems with Applications*, 36(10), 12506-12509. doi. 10.1016/j.eswa.2009.04.032

Doganis, P., Alexandridis, A., Patrinos, P., & Sarimveis, H. (2006). Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. *Journal of Food Engineering*, 75(2), 196-204. doi. 10.1016/j.jfoodeng.2005.03.056

Dutta, G. Jha, P., Laha, A., & Mohan, N. (2006). Artificial neural network models for forecasting stock price index in the Bombay Stock Exchange. *Journal of Emerging Market Finance*, 5(3), 283-295. doi. 10.1177/097265270600500305

Hamid, S.A., Iqbal, Z. (2004). Using neural networks for forecasting volatility of S&P 500 index futures prices. *Journal of Business Research*, 57(10), 1116–1125. doi. 10.1016/S0148-2963(03)00043-2

Hammad, A.A.A., Ali, S.M.A., & Hall, E.L. (2007). Forecasting the Jordanian stock price using artificial neural network. *Intelligent Engineering Systems through Artificial Neural Networks*, Vol 17, Digital Collection of The American Society of Mechanical Engineers. doi. 10.1115/1.802655.paper42

Hanias, M., Curtis, P. & Thalassinos, J. (2007). Prediction with neural networks: the Athens stock exchange price indicator. *European Journal of Economics, Finance and Administrative Sciences*, 9, 21–27.

Hutchinson, J.M., Lo, A.W., & Poggio, T. (1994). A nonparametric approach to pricing and hedging derivative securities via learning networks. *Journal of Finance*, 49(3), 851-889. doi. 10.3386/w4718

Ihaka, R., & Gentleman, R. (1996). A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299-314. doi. 10.2307/1390807

Jaruszewicz, M., & Mandziuk, J. (2004). One day prediction of NIKKEI index considering information from other stock markets. *Proceedings of the International Conference on Artificial Intelligence and Soft Computing*, 3070, 1130–1135. doi. 10.1007/978-3-540-24844-6_177

Kim, K.-J. (2004). Artificial neural networks with feature transformation based on domain knowledge for the prediction of stock index futures. *Intelligent Systems in Accounting, Finance & Management*, 12(3), 167-176. doi. 10.1002/isaf.252

Kimoto, T., Asakawa, K., Yoda, M. & Takeoka, M. (1990). Stock market prediction system with modular neural networks. *Proceedings of the IEEE International Conference on Neural Networks*, San Diego, pp. 1-16, CA, USA. doi. 10.1109/IJCNN.1990.137535

Kunc, M. (2005). Illustrating the competitive dynamics of an industry: the fast-moving consumer goods industry case study. *Proceedings of the 23rd International Conference of the System Dynamics Society*, pp. 1-42, Boston, MA, USA, July, 2005.

Leigh, W., Hightower, R., & Modani, N. (2005). Forecasting the New York Stock Exchange composite index with past price and interest rate on condition of volume spike. *Expert Systems with Applications*, 28(1), 1-8. doi. 10.1016/j.eswa.2004.08.001

Leung, M.T., Daouk, H., & Chen, A.-S. (2000). Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of Forecasting*, 16(2), 173–190. doi. 10.1016/S0169-2070(99)00048-5

Lewandowska, J. (2012). Identification losses in the FMCG1 sector in the light of the European and global researchers. *Proceedings of FIKUSZ'12 Symposium for Young Researchers*, pp. 101-110, Budapest, Hungary, November 2012.

Moshiri, S., & Cameron, N. (2010). Neural network versus econometric models in forecasting inflation. *Journal of Forecasting*, 19(3), 201-217. doi. 10.1002/(SICI)1099-131X(200004)19:3<201::AID-FOR753>3.0.CO;2-4

Mostafa, M. (2010). Forecasting stock exchange movements using neural networks: empirical evidence from Kuwait. *Expert Systems with Application*, 37(9), 6302-6309. doi. 10.1016/j.eswa.2010.02.091

Ning, B., Wu, J., Peng, H., & Zhao, J. (2009). Using chaotic neural network to forecast stock index. *Advances in Neural Networks*, Lecture Notes in Computer Science, Vol.5551, pp. 870–876 Springer-Verlag, Heidelberg, Germany. doi. 10.1007/978-3-642-01507-6_98

Pan, H., Tilakaratne, C., & Yearwood, J. (2005). Predicting the Australian stock market index using neural networks exploiting dynamical swings and intermarket influences. *Journal of Research and Practice in Information Technology*, 37(1), 43–55. doi. 10.1007/978-3-540-89378-3_53

Perez-Rodriguez, J.V., Torra, S., & Andrada-Felix, J. (2005). Star and ANN models: forecasting performance on the Spanish IBEX-35 stock index. *Journal of Empirical Finance*, 12(3), 490–509. doi. 10.1016/j.jempfin.2004.03.001

Rakicevic, Z., & Vujosevic, M. (2015). Focus forecasting in supply chain: the case study of fast moving consumer goods company in Serbia. *Serbian Journal of Management*, 10(1), 3-17. doi. 10.5937/sjm10-7075

Sen J., & Datta Chaudhuri, T. (2016a). Decomposition of time series data of stock markets and its implications for prediction – an application for the Indian auto sector. *Proceedings of the 2nd National Conference on Advances in Business Research and Management Practices (ABRMP'16)*, pp. 15-28, Kolkata, India, January, 2016. doi. 10.13140/RG.2.1.3232.0241

Sen, J., & Datta Chaudhuri, T. (2016b). A framework for predictive analysis of stock market indices – a study of the Indian auto sector. *Calcutta Business School (CBS) Journal of Management Practices*, 2(2), 1-20. doi. 10.13140/RG.2.1.2178.3448

Sen, J., & Datta Chaudhuri, T. (2016c). An alternative framework for time series decomposition and forecasting and its relevance

for portfolio choice: a comparative study of the Indian consumer durable and small cap sectors. *Journal of Economic Library*, 3(2), 303-326. doi. 10.1453/jel.v3i2.787

Sen, J., & Datta Chaudhuri, T. (2016d). An investigation of the structural characteristics of the Indian IT sector and the capital goods sector – an application of the R programming in time series decomposition and forecasting. *Journal of Insurance and Financial Management*, 1(4), 68-132.

Sen, J., & Datta Chaudhuri T. (2016e). Decomposition of time series data to check consistency between fund style and actual fund composition of mutual funds. *Proceedings of the 4th International Conference on Business Analytics and Intelligence (ICBAI 2016)*, Bangalore, India, December 19-21. doi. 10.13140/RG.2.2.14152.93443

Shen, J., Fan, H. & Chang, S. (2007). Stock index prediction based on adaptive training and pruning algorithm. *Advances in Neural Networks*, Lecture Notes in Computer Science, Vol 4492, pp. 457–464, Springer-Verlag, Heidelberg, Germany. doi. 10.1007/978-3-540-72393-6_55

Singhi, A., Jain, N. & Puri, N. (2015). *Re-Imagining FMCG in India*. CII National Summit Report, December 2015, pp, 1-68. Accessed on: May 10, 2017. [Retrieved from].

Thenmozhi, M. (2006). Forecasting stock index numbers using neural networks. *Delhi Business Review*, 7(2), 59-69. Accessed on: May 10, 2017. [Retrieved from].

Tsai, C.-F. & Wang, S.-P. (2009). Stock price forecasting by hybrid machine learning techniques. *Proceedings of International Multi Conference of Engineers and Computer Scientists*, pp. 755-765, Hong Kong, March 2009.

Tseng, K-C., Kwon, O., & Tjung, L.C. (2012). Time series and neural network forecast of daily stock prices. *Investment Management and Financial Innovations*, 9(1), 32-54.

Vayvay, O., Dogan, O., & Ozel, S. (2013). Forecasting techniques in fast moving consumer goods supply chain: a model proposal. *International Journal of Information Technology and Business Management*, 16(1), 118 – 128.

Vriens, A., & Versteijnen, E. (2007). *Forecasting & Planning in the Food Industry*. EyeOn White Paper. Accessed on: May 10, 2017. [Retrieved from].

Wang, W., & Nie, S. (2008). The performance of several combining forecasts for stock index. *International Seminar on Future Information Technology and Management Engineering*, pp. 450- 455, Leicestershire, United Kingdom, November 2008. doi. 10.1109/FITME.2008.42

Wu, Q., Chen, Y., & Liu, Z. (2008). Ensemble model of intelligent paradigms for stock market forecasting. *Proceedings of the 1st International Workshop on Knowledge Discovery and Data Mining*, pp. 205 – 208, Washington, DC, USA, January 2008. doi. 10.1109/WKDD.2008.54

Zhu, X., Wang, H., Xu, L., & Li, H. (2008). Predicting stock index increments by neural networks: the role of trading volume under different horizons. *Expert Systems with Applications*, 34(4), 3043–3054. doi. 10.1016/j.eswa.2007.06.023

**For Cited this Chapter:**

**Sen, J., & Chaudhuri, T.D. (2020). A predictive analysis of the Indian FMCG sector using time series decomposition - based approach, *In* T.D. Chaudhuri (Ed.), *Overview of the Indian Economy: Empirical analysis of financial markets, exchange rate movements and industrial development* (pp.179-219), KSP Books: Istanbul.**

**Overview of the Indian Economy:** Empirical analysis of financial markets, exchange rate movements and industrial development

Editor: **Tamal Datta Chaudhuri**
Calcutta Business School, Diamond Harbor Road, Bishnupur - 743503, 24 South Parganas, West Bengal, India.

ISBN: 978-605-7736-93-2 (e-Book)
KSP Books 2020
© KSP Books 2020

**Copyrights**
Copyright for this Book is retained by the author(s), with first publication rights granted to the Book. This is an open-access Book distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by-nc/4.0 ).

# Tamal Datta Chaudhuri (Edt.)

Dr. Tamal Datta Chaudhuri did his BA in Economics from Presidency College, Calcutta and his MSc in Economics from Calcutta University. He joined Calcutta University as Lecturer. Subsequently he did MA and PhD in Economics from the Johns Hopkins University, USA. He was a visiting fellow in the University of Illinois at Urbana Champaign. He has taught in Indian Statistical Institute, Calcutta, IISWBM Calcutta, IIFT Calcutta, IBS Hyderabad and Kolkata and Army Institute of Management, Calcutta. He is currently Dean of Calcutta Business School. Dr. Datta Chaudhuri has many publications in national and international journals. After completing 15 years in full time academics, Dr. Datta Chaudhuri joined the financial sector and has worked in Industrial Reconstruction Bank of India (IRBI), later named Industrial Investment Bank of India (IIBI), for 23 years. He became Chief General Manager in Charge of IIBI. He has worked in the whole gamut of activities of a financial institution like designing financial instruments for raising finance from the market, project appraisal, long term lending, short term fund deployment, investment etc. He gained expertise in stock market operations and technical analysis during his tenure. He sat on all major decision making committees of the institution including the Board and has actively interacted with the Ministry of Finance in various matters. He was also a member on Boards of various companies as a nominee director of IIBI. Dr. Datta Chaudhuri participates actively in the stock market and uses technical analysis extensively for his decisions. He teaches courses like Microeconomics, Macroeconomics, Financial Management, Stock Market Simulation Game, Concepts of Management, Security Analysis and Portfolio Management, Corporate Finance, Financial Derivatives Management, Project Appraisal & Finance and Risk Management in Banks. He actively worked with Dalmia Securities, Kolkata in their derivatives trading desk. He delivers lectures to students on Derivatives Trading in different forum and also PRIMIA, Kolkata Chapter. He has been an invitee to the Salzburg Seminar, Austria.

KSP Books

e-ISBN
978-605-7736-93-2
KSP Books 2020